THE UNIVERSITY
of EDINBURGH

Enabling stroke and blood pressure research in UK Biobank

Rebecca Woodfield

Doctor of Philosophy

The University of Edinburgh 2017

Declaration

I declare that I composed this thesis by myself, and it is my own original work. The thesis has not been submitted in part or in whole for any other degree or professional qualification except as specified.

Rebecca Woodfield
28th July 2016

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Cathie Sudlow, for providing me with the fantastic opportunity to undertake a PhD. Without her encouragement, I would never have considered a career in research, and I am so grateful to have had this life changing experience. My work has benefitted greatly from her guidance and thoughtful feedback. It was her unwavering enthusiasm which helped me through my fellowship applications, and when times were hard I was especially grateful for her support, patience and flexibility. I am grateful for having had the opportunity to attend numerous conferences and courses, and to contribute to UK Biobank meetings, particularly the outcomes working group meetings in London, where I benefitted from meeting many interesting scientists from across the UK. It has been a pleasure and an inspiration to work with Cathie and to be part of the UK Biobank team.

I would like to thank a number of individuals who have helped me in particular aspects of this research: Dr Ian Grant, statistician at the Information and Services Division, Scotland, for his help with the literature searches, study selection, and data extraction in Chapter 2; Dr Qiuli Zhang, a statistical epidemiologist in Edinburgh, for performing the data analysis in Chapter 6, for her helpful feedback on this chapter, and for providing general statistical/epidemiological advice, particularly as I got to grips with *stata* and with the SAIL primary care databases; Dr Robin Flaig, for her help within UK Biobank, for providing me with useful contact information, and for setting up many beneficial meetings; Professor Peter Rothwell, Oxford University, for his invaluable advice on the potential feasibility of using coded primary care data to capture blood pressure variability (Chapter 8), and for his feedback on some of my early analyses; Dr Mark Atkinson, a statistician at Swansea University, for his help preparing the early Read code lists for ‘stroke’ and ‘blood pressure’, for his support in navigating the SAIL primary care database (particularly with regard to coding BP values), and for his insightful feedback on an early draft of Chapter 8; Dr Spiros Denaxas, for enabling me to access the useful CALIBER website in order to compile my list of Read codes for BP.

I am grateful to the Edinburgh Stroke Research Group members, and members of the Edinburgh Cognitive Ageing and Epidemiology Group, in particular Dr Will Whiteley, Professor Rustam Al-Shahi Salman, Professor Martin Dennis, Professor Peter Sandercock, Dr Susie Shenkin, and Professor Alasdair MacLullich, for their teaching, both clinical and research-orientated, mentorship, career guidance, feedback on numerous drafts of manuscripts/fellowship applications, and for proving to me that clinical-academic medicine is possible! I am particularly grateful to Professor Dennis for encouraging me in my early clinical career, and for inspiring my interest in stroke medicine, and to Dr Shenkin for her support both in the early stages of my PhD, and as I transitioned back to clinical medicine. I would also like to thank my educational and clinical supervisors, Dr Sue Pound, Dr Vicky Tallentire, and Dr Claire Gordon, for being so flexible with my clinical commitments/study leave whilst I've written up this thesis.

I have made some wonderful friends in Edinburgh during this PhD. Many thanks to Dr Kristiina Rannikmae, Dr Neshika Samerasekera, Dr Josephine Tsai, and Dr Peter Foley for their camaraderie, for patiently listening when I needed to talk, for always pointing me in the right direction, and for being a valued source of encouragement and motivation. I couldn't have done this PhD without you!

Finally, and by no means least, a heartfelt thanks to my family and friends who have supported me tirelessly over so many years. Without their kindness, patience, practical help, and understanding, none of this would ever have been possible.

Publications and awards relating to this thesis

Papers in peer reviewed journals relating to this thesis

1. **Woodfield R**, UK Biobank Stroke Outcomes Group, UK Biobank Follow-up and Outcomes Working Group, Sudlow CLM. Accuracy of Patient Self-Report of Stroke: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS ONE* 2015: 10(9): e0137538. doi:10.1371/journal.pone.0137538
2. **Woodfield R**, Grant I, UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group, Sudlow CLM. Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS ONE* 2015: 10(10): e0140533. doi:10.1371/journal.pone.0140533

Papers which I have contributed to during my PhD studies

1. Rannikmäe K, **Woodfield R**, Anderson CS, Charidimou A, Chiewvit P, Greenberg SM, Jeng J-S, Meretoja A, Palm F, Putaala J, Rinkel GJE, Rosand J, Rost NS, Strbian D, Tatlisumak T, Tsai C-F, Wermer MJH, Werring D, Yeh S-J, Al-Shahi Salman R, Sudlow CLM. Reliability of intracerebral haemorrhage classification systems: a systematic review. *International Journal of Stroke* 2016: Apr 18. pii: 1747493016641962. [Epub ahead of print]
2. NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurology* 2015; doi:10.1016/S1474-4422(15)00338-5.
3. Ay H, Arsava EM, Andsberg G, Benner T, Brown RD Jr, Chapman SN, Cole JW, Delavaran H, Dichgans M, Engström G, Giralt-Steinhauer E, Grewal RP, Gwinn K, Jern C, Jimenez-Conde J, Jood K, Katsnelson M, Kissela B, Kittner SJ, Kleindorfer DO, Labovitz DL, Lanfranconi S, Lee JM, Lehm M,

Lemmens R, Levi C, Li L, Lindgren A, Markus HS, McArdle PF, Melander O, Norrving B, Peddareddygari LR, Pedersén A, Pera J, Rannikmäe K, Rexrode KM, Rhodes D,.Rich SS, Roquer J, Rosand J, Rothwell PM, Rundek T, Sacco RL, Schmidt R, Schürks M, Seiler S, Sharma P, Slowik A, Sudlow C, Thijs V, **Woodfield R**, Worrall BB, Meschia JF. Pathogenic ischemic stroke phenotypes in the NINDS stroke genetics network. *Stroke* 2014;45(12):3589-3596.

Conference presentations

1. **Woodfield R**, Rannikmäe K, Grant I, Sudlow C. Reliability and feasibility of ischemic stroke classification systems for large epidemiological studies. Platform presentation at the European Stroke Conference in London in 2013. (Published in *Cerebrovascular Diseases*, 2013: 35:159.)
2. **Woodfield R**, Grant I, Sudlow C, UK Biobank. Accuracy of electronic health record data for ascertainment and sub-classification of stroke outcomes in large epidemiological studies. Poster at the UK Stroke forum in Harrogate in 2013 (Published in *International Journal of Stroke*, 2013: 8:41.)
3. **Woodfield R**, Grant I, Sudlow C, UK Biobank. Accuracy of electronic health record data for ascertainment and sub-classification of stroke outcomes in large epidemiological studies. Platform presentation at the Scottish Health Informatics Platform Conference in St Andrews in 2013.
4. **Woodfield R**, Grant I, Sudlow C. Accuracy of electronic health record data for ascertainment and sub-classification of stroke outcomes in large epidemiological studies. Poster at the European Stroke Conference in Nice in 2014 (Published in *Cerebrovascular Diseases*, 2014: 37 Supplement 1:262.)
5. **Woodfield R**, Sudlow CLM. Accuracy of Patient Self-Report of Stroke. Poster at the European Stroke Conference in Nice in 2014. (Published in *Cerebrovascular Diseases, 2014:* 37 Supplement 1:609)
6. **Woodfield R**, Sudlow C. Can routinely collected primary care data be used to measure visit-to-visit blood pressure variability in large epidemiological studies? Poster at the European Stroke Organisation Conference in Barcelona in 2016. (To be published in *International Journal of Stroke*, 2016)

7. Rannikmäe K, **Woodfield R**, Al-Shahi Salman R, Sudlow CLM on behalf of the ICH Classification Collaboration. Reliability of spontaneous intracerebral haemorrhage classification systems: a systematic review. Poster at the European Stroke Organisation Conference in Glasgow in 2015. (Published in *International Journal of Stroke*, 2015:10 Supplement 2:295.)

Awards

1. **Research funding**: Chest Heart and Stroke Scotland Clinical Research Fellowship (£85,014); UK Biobank; Edinburgh Lothian Health Fund "Stroke Endowments" stipend.
2. **Travel Grants** awarded by: The Guarantors of Brain; UK Biobank

Abstract

Background

Blood pressure is one of the most important modifiable risk factors for stroke. Although the influence of an individual's average blood pressure (BP) on their overall stroke risk is well established, visit-to-visit blood pressure variability (BPV) - variation in blood pressure from one clinic visit to the next- may be an independent risk factor for stroke. The influence of BPV on stroke risk in the general population is not fully understood, nor is it known whether associations with BPV vary by pathological stroke type. Very large prospective studies, including exposure measurements of BP and BPV as well as accurate identification, confirmation and sub-classification of large numbers of stroke cases during follow-up, are needed to test the associations between BP parameters, stroke and its main pathological types.

UK Biobank (UKB) is a very large prospective cohort study of ~500,000 middle aged adults recruited from England, Scotland, and Wales between 2006 and 2010. Participants completed a detailed baseline assessment at recruitment (which included self-report of prior stroke and BP measurement). Follow up for health-related outcomes (including new occurrences of stroke) in UKB relies on linkages to routine coded datasets for hospital admissions, death registrations and primary care data. Coded primary care data could also be used to capture novel exposures, like blood pressure variability (BPV).

In this thesis, I aimed to investigate how large prospective epidemiological studies such as UK Biobank might be used to investigate the influence of BP, and in particular BPV, on stroke and its types and subtypes. I did this through advancing understanding of the identification and characterisation of stroke cases in large prospective studies, and of obtaining measures of BPV from linked primary care data. Specifically, I aimed:

(1) to evaluate the accuracy of patient self-report of stroke, the accuracy of routinely available coded healthcare data for stroke, and the reliability and feasibility of ischaemic stroke classification systems for large epidemiological studies such as UKB;

(2) to identify prevalent and early incident stroke cases in UKB using multiple overlapping sources of coded data, and determine the proportions of cases classified into main pathological types of stroke;

(3) to explore the feasibility of using coded primary care data to obtain measures of BPV in UKB.

Methods

(1) I performed a series of systematic reviews of published data on (i) the accuracy of patient self-report of stroke, (ii) the accuracy for stroke and its main pathological types (ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage) of International Classification of Diseases (ICD) coded hospital admissions and death certificates, and Read coded primary care records, and 3) the inter-rater reliability of ischaemic stroke classification systems.

(2) Informed by this work I identified prevalent and early incident stroke cases in UKB using linked coded hospital and death registration data as well as self-report data. In a sub cohort of participants, I was able to assess the additional role in case identification of linked coded primary care data. I compared the numbers of potential stroke cases ascertained by multiple overlapping combinations of these data and examined the proportions classified into the main pathological stroke types.

(3) Finally, I analysed data from about 10,000 Welsh UKB participants with linked coded primary care data to identify those in whom visit-to-visit BPV could be measured using coded systolic blood pressure values (BP). I explored the association between frequency of visits with coded BP values and: participant characteristics; time between visits; mean BPV; standard deviation of BPV (SD BPV). I also calculated within-individual agreement between coded BP and UKB baseline assessment BP.

Results

(1) From my systematic reviews I found that self-report accuracy was strongly influenced by characteristics of the study population. In populations with low stroke prevalence up to 75% of self-reported strokes were false positives. ICD codes for cerebrovascular diseases had a broad range of accuracy for stroke and

its main pathological types, but appropriately selected, 'stroke specific' ICD codes were consistently >70% accurate when compared to an independent reference standard for stroke. Few studies assessed the accuracy of either primary care data or combinations of data sources for stroke.

The overall inter-observer reliability of ischaemic stroke classification systems ranged from moderate to almost perfect. Study characteristics other than classification system accounted for much of the variation in reliability. Additional features which enhanced reliability included use of clear rules, data abstraction protocols, computerised assignment, and reduced number of subtype categories.

(2) The prevalence of stroke in UK Biobank based on linked ICD coded hospital admissions data and participant self-report was ~1.7%. The majority of these prevalent stroke cases were of 'unspecified' stroke type. Incident strokes captured by ICD codes were mostly hospital admitted cases, but a smaller additional proportion were fatal cases not detected in hospital admissions data. The majority (~89%) of ICD coded incident strokes were a specified pathological type. In the sub-cohort of UKB participants with additional primary care data linkage ~20% of potential incident stroke cases were detected by coded primary care data alone.

(3) Among Welsh UKB participants with linked primary care data, around two thirds had sufficient coded data to estimate visit-to-visit BPV any time before recruitment, and just under half had sufficient coded data to estimate BPV during the 5 years before recruitment. Selecting participants with more visits reduced generalizability, but there was good variability in BPV amongst those selected (standard deviation in BPV range ~5mmHg to ~7mmHg), and reasonable agreement between coded BP and BP recorded at the UKB baseline assessment (intra class correlation coefficient 0.53, 95% CI 0.52 to 0.55).

Conclusions

This work will inform the approaches to stroke outcomes ascertainment and the measurement of a novel exposure, blood pressure variability, in UK Biobank. This will enable future exploration of the associations between blood pressure parameters,

stroke, and its main types and sub-types in UK Biobank. Investigating these associations will improve our understanding of causal pathways for the different pathological types and sub-types of stroke and underpin increasingly targeted strategies to modify BP for stroke prevention.

Contents

Table of Figures

Table of tables

Chapter 1 Introduction

- Stroke is a major cause of death and disability worldwide.
- The main pathological types and subtypes of stroke may differ in their risk factor associations.
- UK Biobank (UKB), a very large prospective cohort study, provides the perfect opportunity to reliably test these associations.
- Blood pressure (BP) is one of the most important modifiable risk factors for stroke. Visit-to-visit BP variability (BPV) - variation in BP from one clinic visit to the next - may confer an additional independent risk of stroke.
- Investigating the associations between BP/BPV and the main stroke types and subtypes will improve our understanding of stroke aetiology and underpin increasingly targeted strategies for stroke prevention.
- My work is divided into three sections:

  A) Systematic reviews: informing the development of accurate and scalable methods for the ascertainment, confirmation and classification of stroke cases in UK Biobank.

  B) Ascertainment of stroke and its main pathological types in UK Biobank: exploring the contribution of multiple overlapping sources of coded healthcare data to the identification of prevalent and early incident stroke.

  C) Measurement of a novel exposure, blood pressure variability: exploring the feasibility of using routinely collected coded primary care data to measure BPV amongst UKB participants.

1.1 Stroke epidemiology

1.1.1 The global burden of stroke

Stroke is the second commonest cause of death worldwide, and a major cause of global disability.(Lozano et al. 2012) In the UK stroke is the most common cause of severe disability amongst adults, and the third commonest cause of death.(Adamson, Beswick and Ebrahi 2004, Murray et al. 2013) The long term consequences of stroke negatively impact patients, their families and society. Half of stroke survivors are left

with physical disability, cognitive impairment, or mood disorder,(Ch'Ng, French and McLean 2008) which can lead to depression, anxiety and social withdrawal amongst patients and their carers.(Murray, Young and Forster 2007) Stroke already accounts for an estimated £9 billion per year cost to the UK economy from hospital treatment, community care, disability payments and loss of income.(Saka, McGuire and Wolfe 2009)

The global burden of stroke has grown in the last 10-20 years due to rising stroke incidence in low and middle income countries, an ageing population, and increased longevity.(Lopez and Mathers 2006, Murray et al. 2012, Feigin et al. 2014) This problem is projected to increase exponentially in the years ahead as the epidemiological shift towards non-communicable diseases continues, stroke survival improves, and patients live longer with disability.(Mukherjee and Patil 2011, Magnusson 2009) Reducing the burden of chronic diseases has therefore become a major global health priority. The World Health Organisation (WHO) has a strategic aim to reduce mortality from non-communicable diseases by 2% by 2015.(Magnusson 2009) Improving primary prevention of stroke, the second largest contributor to this global epidemic, is a key component of this aim.

1.1.2 Epidemiological definitions of stroke

The World Health Organisation (WHO) defines stroke as

> ‘rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours, or leading to death, with no apparent cause other than of vascular origin’.

(Hatano 1976, Aho et al. 1980) Stroke is a form of ‘cerebrovascular disease’, a broader diagnostic term which also includes Transient Ischaemic Attack (TIA), sub-dural and extra-dural haemorrhage, and other vascular conditions affecting the brain.(Whisnant et al. 1990) TIA is a ‘transient’ version of stroke. The distinction between stroke and TIA (WHO definition, above), has traditionally been based on symptom duration. A neurological deficit of presumed vascular origin is ‘stroke’ if symptoms persist more than 24 hours, and ‘TIA’ if symptoms resolve within 24 hours, provided non-vascular diagnoses have been excluded. Although this 24 hour

cut-off was chosen arbitrarily, the 'symptom-based' definition of stroke ensures that the diagnoses (stroke versus TIA) are made consistently. Diagnostic consistency is an important requirement for epidemiological comparisons.(Sudlow and Warlow 1996)

It is increasingly recognised that some TIAs show evidence of infarction on brain imaging.(Ay et al. 2005b) This has led to the development of a new 'tissue-based' definition of stroke, which relies on the detection of brain infarction to diagnose 'stroke', irrespective of symptom duration (<24 hours).(Albers et al. 2002) This definition re-classifies cases of 'TIA with infarction' as stroke. However, the choice and timing of brain imaging is a crucial determinant of the ability to detect infarction in stroke (or TIA), with MRI being more sensitive for early ischaemic change than CT.(Chalela et al. 2007) If the 'tissue-based' definition is used, the number of cases diagnosed as stroke (including TIA with infarction) will depend on the choice and availability of brain imaging, which might vary from one region to another.(Brown, Rudd and McGovern 2003b) This definition is therefore more susceptible to global and regional differences in clinical practice. By contrast the 'symptom-based' definition relies on a competently taken medical history (focused on the nature, onset and duration of symptoms), but not on variable access to imaging or other technology, meaning that it can be applied more consistently, and is equally applicable in resource poor settings, irrespective of access to brain imaging.

1.1.3 Pathological types and subtypes of stroke

Stroke is a heterogeneous clinical syndrome. The World Health Organisation has divided stroke into four pathological types – cerebral infarction (or ischaemic stroke), intracerebral haemorrhage (blood vessel rupture with bleeding within brain tissue), subarachnoid haemorrhage (blood vessel rupture with bleeding within the subarachnoid space), and unspecified type.(Aho et al. 1980) In recent community based studies in the UK, around 80% of strokes are ischaemic, 10% are due to primary intracerebral haemorrhage, 5% subarachnoid haemorrhage, and 5% unspecified.(Rothwell et al. 2004)

The main types of stroke can be further broken down into different anatomical or aetiological subtypes. Around 50% of ischaemic strokes can be attributed to large artery atherosclerosis, 25% to small vessel disease, 20% to cardiac emboli, and 5% to other rare causes.(Warlow et al. 2003) Intracerebral haemorrhage can be classified into supra tentorial (lobar and non-lobar haemorrhage), cerebellar haemorrhage, brainstem haemorrhage and intra-ventricular haemorrhage, and subarachnoid haemorrhage can be classified into aneurysmal and non-aneurysmal subtypes (Figure 1.1).

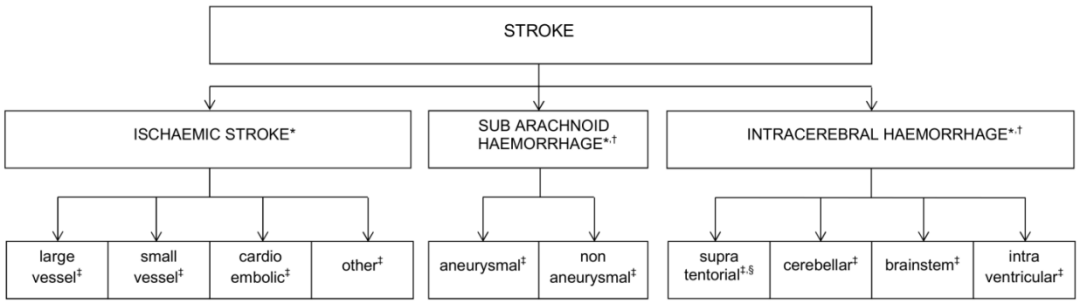

Figure 1.1 Stroke classification. The main pathological types and sub-types of stroke.

*The main pathological types of stroke

†Subarachnoid haemorrhage and intracerebral haemorrhage together make up 'haemorrhagic stroke'.

‡Pathological sub-types of stroke. There is also an 'uncertain' subtype (not shown).

§Supratentorial haemorrhage can be further divided into lobar and non-lobar haemorrhage.

The main types of stroke share common risk factors,(Warlow et al. 2003) but evidence suggests that some of these risk factors may associate more strongly with haemorrhagic or ischaemic stroke types.(Asia Pacific Cohort Studies Collaboration 2005, O'Donnell et al. 2010)

Furthermore, there is evidence from case-control studies for differences between ischaemic stroke subtypes in their associations with vascular risk factors, genetic polymorphisms, blood brain barrier integrity, and retinal vasculature morphology,(Jackson et al. 2010, Gschwendtner et al. 2009, Bellenguez et al. 2012, Wardlaw et al. 2009, Doubal et al. 2009, Lindley et al. 2009) and evidence of differences between these stroke subtypes in their prognosis with respect to vascular

outcomes including myocardial infarction and recurrent stroke.(Jackson et al. 2009) There is also evidence for different genetic and non-genetic risk factors for lobar versus non-lobar intracerebral haemorrhage.(Jackson and Sudlow 2006, Biffi et al. 2010) These differences may point to different mechanistic pathways for the different stroke types and subtypes. If better understood, these different pathways may provide potential new targets for drug development, and in the long term, a more stratified approach to the prevention of stroke.

1.2 The association between blood pressure and stroke

1.2.1 Mean blood pressure and stroke

Blood pressure is one of the most important modifiable risk factors for stroke.(Lawes, Vander Hoorn and Rodgers 2008) There is a steep log-linear relationship with stroke such that sustained reductions in usual BP of 20mmHg systolic/10mmHg diastolic approximately halve total stroke risk in middle to old age.(Lewington et al. 2002) The influence of BP on stroke risk in general is well established, but there remain uncertainties about whether this differs between the pathological types and subtypes, and if so in what way. Stronger associations have been suggested between BP and intracerebral haemorrhage versus ischaemic stroke,(Asia Pacific Cohort Studies Collaboration 2005) lacunar ischaemic stroke versus other ischaemic subtypes,(Jackson et al. 2010) and lobar versus non-lobar intracerebral haemorrhage,(Jackson and Sudlow 2006) but robust evidence for such differences is limited.

1.2.2 Blood pressure variability and stroke

While the most comprehensive evidence to date focuses on associations with usual (long-term average) systolic, diastolic and mean BP,(Lewington et al. 2002, Asia Pacific Cohort Studies Collaboration 2005) recent work also suggests independent effects of episodic hypertension and medium to long-term BP fluctuations.(Rothwell et al. 2010b) Evidence that BP variability independently influences stroke risk may have important implications for BP monitoring and treatment. For example, some anti-hypertensive medications reduce BP variability more than others.(Rothwell et al. 2010a) Better understanding of the associations between various BP parameters and

stroke, its types and subtypes, should both improve our understanding of type/subtype specific causal pathways and underpin increasingly targeted strategies to modify BP for stroke prevention.

1.2.3 Limitations of previous studies

Previous studies which examined the associations between usual (mean) blood pressure and the main types and subtypes of stroke have had a number of limitations. Retrospective studies (recruiting patients with stroke +/- controls) had relatively crude measures of BP and other exposures, and were prone to recall and control selection bias.(Jackson and Sudlow 2005, Jackson et al. 2010, O'Donnell et al. 2010) Individual prospective studies have generally been too small to address associations between BP and stroke reliably. Collaborative meta-analyses of prospective studies provided adequate sample sizes, but had other limitations including pooling of heterogeneous data, outcome selection bias, and incomplete or inaccurate classification of stroke types/subtypes.(Lewington et al. 2002, Asia Pacific Cohort Studies Collaboration 2005) Most stroke outcomes were ascertained from death certificates and/or hospital discharge data, meaning that fatal strokes were overrepresented, and non-hospitalised non-fatal stroke cases were not included. This may have led to over selection of haemorrhagic stroke cases, which are more likely to be fatal than ischaemic strokes,(Andersen et al. 2009) and under selection of lacunar ischaemic strokes, which are less likely to be hospitalised than non-lacunar ischaemic stroke.(Schulz and Rothwell 2003, Bejot et al. 2013, Giroud et al. 1997) Furthermore, information on main pathological type was available for only around half of the stroke outcomes in these studies, and further stroke classification into subtypes was not possible.

1.3 Stroke in clinical practice

1.3.1 Diagnosis of stroke

There is no ‘gold standard’ test for the diagnosis of stroke. The most accurate approach requires expert assessment of the presenting clinical syndrome aided by appropriately timed brain imaging.(Warlow et al. 2003) Brain imaging (CT or MRI) is useful to exclude ‘stroke mimics’, but such imaging is not uniformly available, and requires appropriate timing and expert interpretation. Even if every stroke patient had

brain imaging, CT or MRI would fail to demonstrate a relevant stroke lesion in some of the milder cases of stroke. Clinical assessment is therefore the most important aspect of stroke diagnosis. However, agreement between physicians is imperfect,(Kessler et al. 1991) and around 30% of potential strokes admitted to hospital leave with a non-cerebrovascular diagnosis.(Hand et al. 2006b) Even amongst experts up to 5% of initial stroke diagnoses are incorrect.(Ricci, Celani and Righetti 1994)

1.3.2 Diagnosis of TIA

The diagnosis of TIA is even more challenging. In the majority of cases the symptoms and signs of TIA have already resolved by the time the patient attends for assessment. Diagnosis depends almost entirely on the ability of the physician to elicit an accurate history. A wide array of conditions mimic TIA including focal seizure, migraine, labyrinthitis, and functional symptoms.(Nadarajan et al. 2014) It is therefore not surprising that the reliability of diagnosis is poor, even amongst experts.(Castle et al. 2010)

For some research studies it may not be important to distinguish between TIA and stroke. TIA and stroke share many common risk factors and common secondary prevention strategies. In the acute phase (first few hours) of presentation, patients with TIA or stroke are often grouped together with the diagnosis ‘acute brain attack’, before a formal diagnosis is made. However if diagnostic accuracy is important for a study it may be sensible to separate stroke from TIA. This is because stroke diagnosis, albeit imperfect, is more accurate than TIA diagnosis, and less likely to introduce false positive diagnoses. If the aim is to maximise positive predictive value (see 1.4.3), then inclusion of TIA risks inclusion of a wide variety of potential TIA mimics. The WHO definition of stroke is also preferable, because this is less susceptible to the availability and timing of investigation, and is more consistent when applied across multiple settings.

1.3.3 Pathways of care in the UK

In the UK, ~150,000 individuals have a stroke each year.(Townsend et al. 2012) Between 60% to 90% of stroke patients are admitted to hospital,(Schulz and

Rothwell 2003), but this is likely to vary by region (and potentially by stroke pathology).

In England, Scotland, and Wales ~80% of hospital admitted patients are admitted to expert led stroke units,(SSCA 2015, SSNAP 2015) and the remainder are treated on general medical wards with or without specialist input. The majority of hospitalised patients receive brain imaging (CT or MRI) but clinical practice varies by region. National targets require 90% of patients admitted to hospital with stroke to have had brain imaging within 24hours of admission.(SSNAP 2015, SSCA 2015) Investigation of patients is likely to be more complete amongst patients treated in specialist units.

Non-hospitalised strokes include fatal cases (not reaching hospital prior to death), and mild cases, not severe enough to warrant hospital admission and often with rapid and complete or near complete recovery. Non-fatal, non-hospitalised stroke patients have access to expert led diagnosis and investigation through referral (by general practitioners or acute medical physicians) to dedicated stroke outpatient clinics. However, depending on local referral patterns and patient specific factors, some of these non-hospitalised patients might never receive an 'expert' diagnosis, instead being treated by their GP (perhaps in a residential or care home setting), or by physicians in hospital without referral to acute stroke services. Depending on the structure of local stroke services, there is therefore potential for regional variation in the accuracy of 'stroke diagnoses.

1.4 UK Biobank

1.4.1 Background and aims

UK Biobank is a very large population-based prospective cohort study which was designed to allow the assessment of the relevance of many different exposures to many different outcomes.(Collins 2012) Common conditions, like stroke, have numerous genetic and environmental determinants.(Manolio, Bailey-Wilson and Collins 2006) Very large studies, generating large numbers of outcomes for adequate statistical power, are required to study the modest individual effects of these determinants and the complex interactions between them.(Burton et al. 2009)

UK Biobank fulfils the key requirements of an ideal study of disease aetiology.(Grimes and Schulz 2002) These include:

- prospective study design, with accurate measurement of the relevant exposures and covariates before the occurrence of outcomes, to avoid reverse causality;
- selection of non-disease controls from the same population as the cases, to avoid control selection bias;
- repeat measures of baseline parameters and covariates in a subset of the cohort during follow-up, to correct for regression dilution bias, where the strength of the association is underestimated;
- very large sample size (hundreds of thousands of participants), accruing a large enough number of disease outcomes (thousands overall) for sufficient statistical power;
- accurate confirmation and detailed sub-classification of disease outcomes.

1.4.2 Participant recruitment

Around 500,000 participants aged 40-69 years were recruited to UK Biobank between April 2007 and July 2011.(Sudlow et al. 2015) Invitations were sent by post to potential participants inviting them to attend one of 22 assessment centres in England, Scotland and Wales. A detailed baseline assessment was conducted at the recruitment visit. This included a touch-screen questionnaire, a brief nurse-led interview, physical examination, and collection of blood, urine and saliva to be stored for future genetic and biochemical analyses. Around 20,000 to 25,000 participants are recalled every few years to repeat this assessment, enabling adjustment for measurement error and variation in baseline parameters over time.

Table 1.1 shows the baseline characteristics of the UK Biobank population, including prevalent conditions at recruitment.(http://biobank.ctsu.ox.ac.uk/crystal/) In general, participants were better educated and less socioeconomically deprived than the rest of the UK (healthy cohort-effect) but all socioeconomic strata were represented. Participants were 46% male, 85% urban, 94.5% white (the remaining 5.5% reflecting ethnic mix for the UK), ~60% in paid employment, 57% aged 40-49, and 43% aged 50-69.

Table 1.1 Baseline characteristics of the UKB population

Characteristic(s)	UKB participants (n & %)[*]
Gender	
Male	229,171 (46)
Female	273,461 (54)
Age range	
40-49	117,897 (24)
50-59	167,161 (33)
60-69	217,465 (43)
Country	
England	444,000 (89)
Scotland	36,000 (7)
Wales	21,000 (4)
Prevalent conditions at recruitment[†]	
Diabetes	26,000 (5)
Myocardial infarction	12,000 (2)
Pulmonary Disease	12,000 (2)
Stroke	7,000 (1)
Rheumatoid arthritis	6,000 (1)
Smoking status[†]	
Current smoker	52, 974 (11)
Ex-smoker	173,066 (34)
Never smoked	273,545 (54)
Ethnic background[†]	
White	472,371 (94)
Asian	9,548 (1.9)
Black	7,538 (1.5)
Chinese	1,508 (0.3)
Mixed	3,015 (0.6)
Other	4,523 (0.9)
Employment status[†,‡]	
Paid employment	301,199 (60)
Retired	196,284 (39)
Unemployed	9,823 (2)
Disability allowance	20,861 (4)
Other[§]	52,636 (10)
Education/qualifications[†,‡]	
University degree	176,216 (35)
A level/equivalent	141,935 (28)
O level/equivalent	242,106 (48)
Vocational qualification(s)	151,263 (30)
None/no answer	95,533 (19)
Townsend deprivation index	
Mean -1.3 (standard deviation 3, range -6.3 to 11)	

[*]Amongst 502, 523 recruited participants, rounded to the nearest integer

[†]Self-reported at the UKB baseline assessment and confirmed by trained interviewer, rounded to the nearest 1,000.

[‡]Participants were able to select more than one category.

[§]Unpaid voluntary work, family/childcare, student, or no answer.

While UK Biobank participants are not representative of the general population (and hence cannot be used to provide representative disease prevalence and incidence rates), the large sample size and heterogeneity of exposure measures included, allow for valid scientific inferences of associations between exposures and health outcomes that are generalizable to the wider population.(http://www.ukbioabank.ac.uk/wp-content/uploads/2017/03/access-maters-representativeness-1.pdf)

1.4.3 Follow-up strategy

UK Biobank (UKB) participants have consented to follow-up of all health-related outcomes through linkages to National Health Service data including electronic hospital, death certificate, and primary care data, and detailed medical record data. By the end of 2017 around 5,000 incident stroke cases are expected to have occurred in UKB, rising to 9,000 cases by 2022 (estimates based on UK age- and sex-specific rates, adjusted for potential losses to follow-up and the healthy cohort effect).(Sudlow et al. 2015)

The challenge UK Biobank faces is the accurate identification, confirmation and sub-classification of these many thousands of cases of a wide range of diseases, including stroke, from within a cohort of half a million individuals. Firstly, the approach needs to be feasible, meaning that where possible, involvement of expert researchers is minimised, costs are reduced, and the availability of resources is taken into account (given that follow-up is based on data collected during routine NHS care). Secondly, the approach needs to be scalable and geographically generalizable, and therefore must operate consistently across England, Scotland and Wales, despite potential variation in clinical practice, data collection processes and/or data structure between these three countries. Finally, the approach needs to be future proof, meaning that it can adapt to future changes in diagnostic practice, and/or NHS structure. These considerations need to be balanced with diagnostic accuracy, recognising that misdiagnosis/misclassification of disease outcomes will reduce statistical power to detect differences in exposure-outcome relationships between them.(Jaffar et al. 2003a, Choi et al. 2002a)

With these challenges in mind, UK Biobank has proposed a staged approach to outcomes adjudication. Using stroke as an exemplar disease, the first stage will use cohort-wide linkages to coded electronic health record data to identify as many potential cases as possible. This stage should have reasonably high sensitivity, which is the proportion of 'true stroke' cases in the UKB population which are detected. The second stage will confirm which of the potential cases already identified, are 'true stroke cases'. This stage should maximise Positive Predictive Value (PPV), which is the proportion of potential strokes (identified by coded data or other sources) which are confirmed cases. This stage might use combinations of coded data or coded data plus brain imaging to confirm stroke versus non-stroke, or identify the main stroke type. The final stage will involve sub-classification of confirmed stroke cases into their main pathological types and subtypes. This is likely to require detailed review of individual hospital case records and primary care records. Progression through each of these stages results in fewer numbers of cases (fewer confirmed cases need to be classified than the number of potential cases requiring confirmation), but more detailed assessment is required (more time and information is required to classify into ischaemic or haemorrhagic subtypes than to confirm stroke versus non-stroke).

The disease outcomes identified in UKB will be made available to future researchers, allowing them to construct nested case-control or case-cohort studies, in order to answer specific research questions. The emphasis in disease outcomes ascertainment, confirmation and sub-classification is therefore on detecting as large a number of cases as possible without compromising positive predictive value. Maintaining sensitivity will optimise the number of stroke cases available for future analyses, and will preserve statistical power. Maximising PPV (minimising false positive cases), will increase statistical power, because misclassification of outcomes reduces statistical power to detect exposure-outcome relationships. Some false negatives can be tolerated, because these will be diluted in the larger control population, with much more limited impact on statistical power.

1.5 Summary and aims

UKB provides the perfect opportunity to overcome the limitations of previous epidemiological studies, and to better understand the associations between BP parameters (mean BP, and BP variability) and the main types and sub-types of stroke. My aims in this thesis were, firstly, to investigate and develop accurate, scalable methods for the ascertainment, confirmation and sub-classification of stroke cases in large prospective studies, like UK Biobank. Secondly, I aimed to explore methods to identify a novel exposure, blood pressure variability, in large populations, using primary care coded data. The thesis is divided into three main sections.

Section A. Systematic reviews: using evidence to inform the strategy for ascertainment, confirmation and sub-classification of strokes in large prospective studies.

I present the results of two systematic reviews which investigate what is known from existing studies about the accuracy of routinely available coded healthcare data (Chapter 2), and patient self-report (Chapter 3) for stroke and its main types. I focus on assessment of sensitivity and PPV (see staged approach to outcomes adjudication, 1.4.3). Finally, I present results of a systematic review investigating the reliability and feasibility of ischaemic stroke classification systems for large epidemiological studies (Chapter 4)

Section B. Cross referencing analyses: investigating routinely available sources of stroke cases in UK Biobank.

I present the process of creating a 'stroke case definition' using primary care coded data (Chapter 5). I then present cross-referencing analyses which explore the contribution of various data sources (hospital data, death certificate data, primary care data and self-report) to stroke diagnoses in a subset of Welsh UK Biobank participants (Chapter 6).

Section C. Assessing the potential of coded primary care data to capture blood pressure variability in UK Biobank.

I present a review of the evidence from existing published studies for a potential association between blood pressure variability and risk of stroke (Chapter 7). I then go on to explore the feasibility of using coded primary care data to measure the novel exposure of blood pressure variability among UK Biobank participants (Chapter 8).

Ultimately, this work should enable future analyses of the associations between a wide range of exposures (or risk factors) - in particular, blood pressure parameters (including blood pressure variability) - and the different types and subtypes of stroke in UK Biobank.

Section A: Systematic reviews

Chapter 2 Accuracy of coded healthcare data for identifying stroke cases in large epidemiological studies: a systematic review.

- Follow-up in UK Biobank is initially through linkages to coded national health care data.
- Our knowledge of the accuracy of these data for identifying and confirming stroke cases is incomplete.
- The ideal follow-up strategy in UK Biobank would identify as many true stroke cases as possible (adequate sensitivity) without introducing too many false positives (maximising positive predictive value, PPV).
- In this chapter I present the results of a systematic review of the accuracy of coded healthcare data for stroke and its main pathological types.
- I used a comprehensive search strategy, critically appraised study quality, and assessed the influence of the codes selected on sensitivity and PPV.
- I conclude that appropriately selected 'stroke specific' codes are sufficiently accurate (high PPV) for identifying stroke cases when further confirmation steps are not available. There were insufficient data on the accuracy of primary care coded data or multiple overlapping data sources for stroke.

2.1 Introduction

Long-term follow-up in UK Biobank is chiefly through linkages to coded national health care data.(Sudlow et al. 2015) For health-related outcomes such as stroke, UK Biobank aims to maximise statistical power to detect genuine associations in nested case-control or case-cohort studies. This requires a strategy that identifies cases representative of the spectrum of the disease being studied with adequate sensitivity, and that maximises positive predictive value (PPV, the proportion of cases that are true positives). UK Biobank aims to fulfil these requirements by using multiple sources of coded data (primary care, hospital and death certificate data) to ascertain possible stroke cases, and then to implement algorithms, using combinations of coded data, supplemented where required by more detailed medical record review, to

confirm and sub-classify cases of stroke. An important first step in developing such algorithms is to understand the accuracy of the coded data sources.

2.1.1 Routinely collected and coded national healthcare data

In most countries, including the UK, hospital admissions and death certificates are coded using the International Classification of Diseases (ICD).(http://apps.who.int/classifications/icd10/browse/2010/en , http://www.cdc.gov/nchs/icd.htm , http://www.icd9data.com/) The primary ICD code identifies the main condition treated during a hospital admission, or the underlying cause of death. Secondary codes record additional diagnoses relevant to an admission, or contributing to death. Figure 2.1 shows ICD codes for cerebrovascular diseases. The shaded boxes show which ICD codes most closely match the World Health Organisation (WHO) definition of stroke (Hatano 1976) or of one of its three main pathological types: ischaemic stroke, intracerebral haemorrhage (ICH), and subarachnoid haemorrhage (SAH).

Although not all of these codes represent a diagnosis of the clinical syndrome of stroke, many studies which have looked at determinants of stroke using linked ICD-coded datasets have included all cerebrovascular disease codes in the relevant ICD coding chapter, implicitly assuming that they are all codes for stroke. Over the last 10 years, European countries have switched from ICD-9 to ICD-10, while North America uses ICD-9-CM (a clinically-modified version of ICD-9). Primary care data in the UK are coded by general practitioners using the Read coding system, which encodes diagnoses, symptoms, signs, procedures, prescriptions and other administrative data.(Chisholm 1990, Stuart-Buttle et al. 1996)

Figure 2.1a

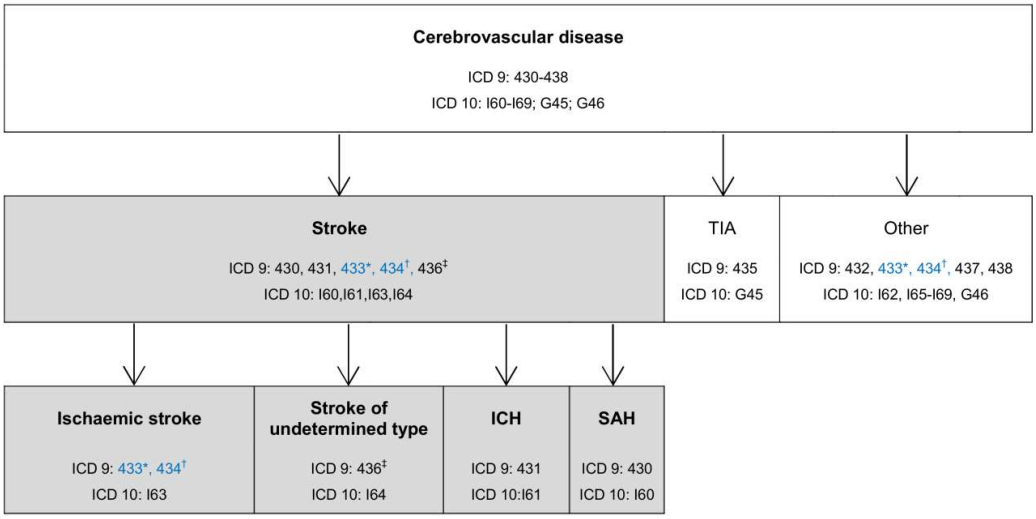

Figure 2.1b

ICD-10	ICD-9	Code description
I60	430	Subarachnoid haemorrhage (SAH)
I61	431	Intracerebral haemorrhage (ICH)
I62	432	Other, non-traumatic intracranial haemorrhage
I63	433*, 434†	Cerebral infarction¶
I64	436‡	Stroke, not specified as haemorrhage or infarction
I65	433*	Occlusion/stenosis of pre-cerebral arteries without infarction
I66	434†	Occlusion/stenosis of cerebral arteries without infarction
I67	437	Other cerebrovascular diseases
I68	-	Cerebrovascular diseases in disorders classified elsewhere
I69	438	Sequelae of cerebrovascular disease
G45	435	Transient ischaemic attack (TIA)
G46§	-	Vascular syndromes of the brain in cerebrovascular disease

Figure 2.1 International Classification of Diseases (ICD) codes for cerebrovascular disease.

*433: occlusion/stenosis of pre-cerebral arteries *with or without infarction.*

†434: thrombosis/embolism of cerebral arteries *with or without infarction.*

Codes in blue text denote ICD-9 codes which most closely represent stroke when subdivided using additional coding available in the clinically modified version of ICD-9 (ICD-9-CM). In ICD-9-CM, '*with infarction*' (433.x1, 434.x1) is distinguished from '*without infarction*' (433.x0, 434.x0).

‡436: acute, ill-defined cerebrovascular disease

¶a pathological term for ischaemic stroke

§G46*:* not a diagnostic code; may be used for the presenting symptoms of either stroke or TIA.

2.2 Aims

To inform approaches to ascertainment, confirmation and sub-classification of stroke in UK Biobank and other large epidemiological studies, I performed a systematic review of published studies of the accuracy of coded health record data for stroke and its main pathological types (ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage). I chose not to include transient ischaemic attacks (TIAs), which are clinically harder to diagnose accurately, with poor agreement even amongst experts,(Ferro et al. 1996b) and of substantially less public health impact than strokes. I used the traditional, epidemiological 'symptom-based' definition of stroke (symptom duration >24 hours) to distinguish stroke from TIA.(Hatano 1976) The more recent, alternative 'tissue-based' definition relies on the presence of brain infarction to diagnose stroke, irrespective of symptom duration (<24hours).(Albers et al. 2002) Accurate diagnosis of brain infarction depends on the availability, choice, and timing of brain imaging, which may vary between different centres. (Brown, Rudd and McGovern 2003a) I chose to use the 'symptom-based' definition to maximise comparability between different studies.

2.3 Methods

The study protocol is displayed in Appendix 2.7

2.3.1 Search strategy

I searched Medline and Embase from 1990 to November 2013 for studies which compared electronic health record data cerebrovascular disease codes against a second data source for stroke or its main types. I used a combination of medical subject heading and text word terms for 'cerebrovascular disease', 'stroke', 'medical records', 'clinical coding', and 'validation studies' (Appendix 2.7.2). I also reviewed bibliographies of included primary studies and relevant reviews, as well as lists of publications from the Clinical Practice Research Datalink(http://www.cprd.com/Bibliography/Researchpapers.asp) and The Health Improvement Network(http://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/publications) websites to identify additional studies evaluating accuracy of primary care data.

2.3.2 Eligibility criteria

Included studies had to have assessed International Classification of Diseases (ICD) or Read coded events against a reference standard data source for stroke, or of one or more of its three major pathological types,(Figure 2.1) defined according to WHO or equivalent definitions.(Hatano 1976) Studies had to report which codes were validated and either their positive predictive value (PPV) or data from which it could be calculated. I excluded studies with less than 50 coded events (since these would have limited precision) and studies in highly selected populations at increased risk (e.g. those with vascular risk factors or known vascular disease) of stroke because of the influence of stroke prevalence on PPV. An epidemiologist colleague (Dr Ian Grant) reviewed all titles and abstracts to select potentially relevant studies for inclusion. I independently reviewed a subset of 10% of these titles and abstracts. We both independently reviewed full texts of potentially relevant studies and selected studies for inclusion. Any areas of uncertainty from this two phase study selection process were discussed and resolved with my supervisor, Professor Cathie Sudlow.

2.3.3 Data extraction and analysis

I extracted and tabulated information from each included study on: first author and publication year; geographic setting (country); age (mean and/or range) of included cases (coded events); data source (hospital, death certificates, primary care); coding system and version; codes used to identify cases; diagnostic position of these codes in the electronic health record (primary versus secondary); number of cases (coded events) compared against the reference standard; reference standard used; PPV and, where reported or calculable, sensitivity, specificity, and negative predictive value (NPV) of codes. I only extracted sensitivity, specificity and NPV values where the reference standard was a population-based stroke register which had clearly aimed to include all stroke cases in the population under study.

I assessed study level quality with a modified version of the Quality Assessment of Diagnostic Studies tool (QUADAS-2),(Whiting et al. 2011) adapted from a recent systematic review of the validity of myocardial infarction diagnoses in administrative databases.(McCormick et al. 2014) I used this to assess reporting quality, generalisability to the UK population (because I sought to recommend codes for UK Biobank), and risk of bias. The study protocol (Appendix 2.7.12) provides a detailed list of questions and scoring methods. An overall quality score (0-14) was derived by combining scores for reporting quality, generalisability, and low risk of bias. I did not exclude studies on the basis of quality assessments.

I calculated 95% confidence intervals for PPV, sensitivity, specificity and NPV values in *Stata* (version 12) using the Wilson method for binomial proportions.(Brown, Cai and DasGupta 2001) For stroke and each of its main pathological types, I assessed the influence on PPV (and, where available, sensitivity) of the codes used to identify stroke cases, and of other study characteristics, using visual inspection of tabulated data and forest plots, and making within-study comparisons where possible to minimise bias. I did not undertake formal meta-analyses or meta-regression because of the substantial heterogeneity between studies in their settings, methods and reporting.

2.4 Results

2.4.1 Studies identified

A total of 39 studies fulfilled my inclusion criteria (Figure 2.2). Of these, 37 were of ICD-coded hospital data, death certificates, or both.(Ives et al. 1995, Lakshminarayan et al. 2009, Leibson et al. 1994, Reker et al. 2001, Rosamond et al. 1999, Roumie et al. 2008, Derby et al. 2000, Derby et al. 2001, Liu et al. 1999, Mayo et al. 1993, Klatsky et al. 2005, Tirschwell and Longstreth 2002, Wahl et al. 2010, Goldstein 1998, Benesch et al. 1997, Johnsen et al. 2002, Appelros and Terent 2011, Krarup et al. 2007, Tolonen et al. 2007, Ellekjær et al. 1999, Leone et al. 2004, Stegmayr and Asplund 1992, Spolaore et al. 2005, Haesebaert et al. 2013, Aboa-Eboule et al. 2013, Palmieri et al. 2007a, Sinha et al. 2008, Wright et al. 2012, Davenport, Dennis and Warlow 1996, Mant, Mant and Winner 1997, Barer 1996, Panayiotou et al. 1993, Hasan, Meara and Bhowmick 1995, Kirkman et al. 2009, Koster et al. 2013a, Harriss et al. 2011, Rinaldi et al. 2003) Only two were of Read coded primary care data.(Ruigomez, Martin-Merino and Garcia Rodriguez 2010, Gaist et al. 2013)

2.4.2 Characteristics of studies of hospital and death certificate (ICD-coded) data

Study characteristics are displayed in Table 2.1. The 37 studies were from North America, Europe, (eight UK based), and Australia. They assessed ICD code versions 9, 10, or both. Most studies used hospital data only, but one was of death certificates only, and six used both. The majority of studies sought cases of stroke, 14 sought ischaemic stroke, five haemorrhagic stroke (ICH or SAH), four ICH, and four SAH. The range of codes used varied widely. To identify stroke cases, the largest number of studies used the whole range of cerebrovascular disease codes (either with or without codes for transient ischaemic attack [TIA]), but others used stroke-specific codes. Several others used various miscellaneous groups of cerebrovascular codes to identify stroke cases, while a further four did not include SAH in their definition of stroke and so excluded SAH codes. The diagnostic position of codes was recorded by 31 studies, of which 11 used the primary position alone. Reference standards were either review (generally by a specialist physician) of the hospital or primary care records or a hospital discharge summary, or comparison with a population- or

hospital-based stroke register. Stroke register cases were identified using administrative data (generally multiple overlapping sources), with 'hot pursuit' and confirmation by expert medical record review

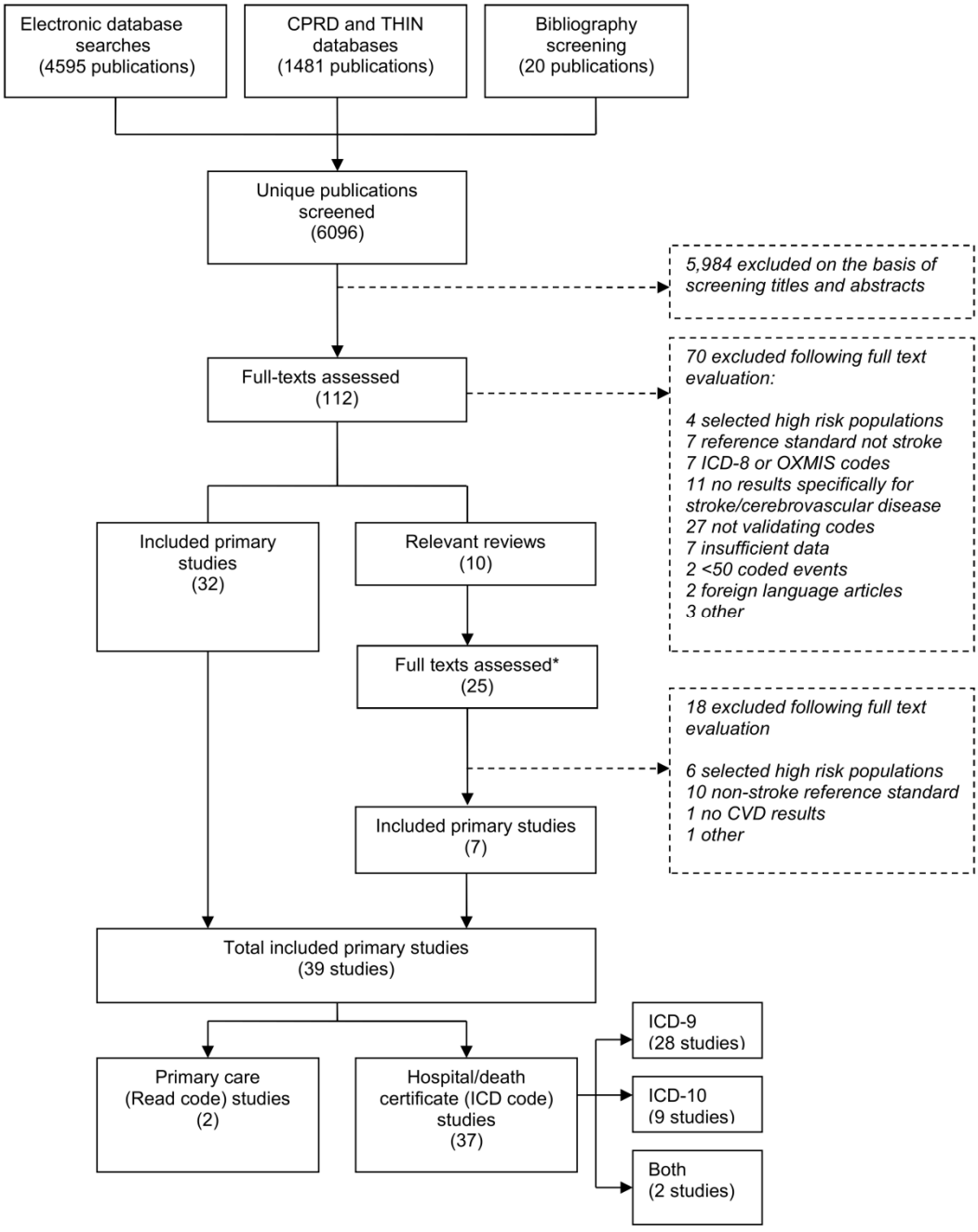

Figure 2.2 Selection of studies

*****Additional studies identified from review articles.

***Table 2.1 Characteristics of studies validating ICD codes from hospital and death certificate data for stroke and its pathological types.*[*]**

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Studies validating codes for stroke											
Mayo 1993	Canada	-	430-434, 436, 437	9	H	96	P	Medical Record	72	76 (67 to 83)	8
Liu 1999	Canada	-	430-438	9	H	862	P/S	Medical Record[¶]	487	56 (53 to 60)	8
			430-438			621	P		417	67 (63 to 71)	
			430-438			327	P/S		151	46 (41 to 52)	
			430-438			213	P		129	61 (54 to 67)	
Leibson 1994	US	-	430-438	9	H	377	P	Population Register	225	60 (55 to 65)	11
			430-438			462	P/S		249	54 (49 to 58)	
			430-438			-	P	Hospital Register	239	-	
			430-438			-	P/S		290	-	
Rosamond 1999	US	45-64	430-438	9-CM	H	1058	-	Medical Record[**]	326	31 (28 to 34)	12
			430-434			526			234	44 (40 to 49)	
Klatsky 2005	US	-	430-438	9	H	3239	P	Medical Record[**]	2494	77 (76 to 78)	10

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Reker 2001	US	-	430-438	9-CM	H	671	-	Medical Record[**]	279	42 (38 to 45)	6
			430, 431, 433, 434, 436			334	-		198	59 (54 to 64)	
			430, 431, 432, 434, 436 OR 430-438[††]			491	P/S	Inpatient register	254	52 (47 to 56)	
			431.x, 433.x1, 434.x1			200	P/S		150	75 (69 to 80)	
Leone 2004	Italy	-	430-438	9	H	1017	P/S	Inpatient Register	609	60 (57 to 63)	11
			430-438			833	P		550	66 (63 to 69)	
			430, 431, 434, 436			411	P/S		371	90 (87 to 93)	
			430, 431, 434, 436			375	P		353	94 (91 to 96)	
Sporalore 2005	Italy	-	430-434, 436-438	9	H	3619	P/S	Medical Record	1296	36 (34 to 37)	9
			430-434, 436-438			2174	P		1021	47[‡‡] (45 to 49)	
Palmieri 2007	Italy		342, 430-438	9	H+D	2793	P/S	Population register	1173	42 (40 to 44)	8
Stegmayr 1992	Sweden	25-74	430-438	9	D	899	P/S	Population Register	812	90 (88 to 92)	12
			430-438		H	5101	P/S		3492	69 (67 to 70)	
Ellekjaer 1999	Norway	≥ 15	430-438	9	H	759	P/S	Population Register	369	49 (45 to 52)	9
			430,431,434, 436			508	P/S		347	68 (64 to 72)	
Panayiotou 1993	UK	25-100	430-438	9	H	117	P	Inpatient Register	94	80 (72 to 87)	8
Hasan 1996	UK	60-94	430-438	9	H	166	P	Medical Record	113	68 (61 to 75)	9

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Harriss 2010	Australia	40-69	430-438, I60-I69	9 +10	D	119	P	Medical Record	72	61 (52 to 69)	12
			430-431, 433-434, I60, I61, I63, I690, I691, I693			61	P		54	89 (78 to 94)	
			430-438, I60-I69			-	P/S		-	-	
Johnsen 2002	Denmark	50-64	I60-I69 and G45	10	H	565	P/S	Medical Record	325	58	10
			I60, I61, I63, I64			378	P/S		299	79 (75 to 83)	
Krarup 2007	Denmark	-	I60-I69, G45	10	H	236	-	Medical Record	153[‡‡]	65[‡‡] (59 to 71)	9
			I60, I61, I63, I64			164	-		136[‡‡]	83[‡‡] (76 to 88)	
Sinha 2008	UK	40-79	I60-I69	10	H + D	250	P/S	Medical Record	191	76 (71 to 81)	11
Roumie 2008	US	50-84	430, 431, 433.x1, 434.x1, 436[§§]	9-CM	H	231	P/S	Medical Record[**]	205	89 (84 to 92)	9
			430, 431, 433.x1, 434.x1, 436	9-CM		203	P	Medical Record[**]	196	97 (93 to 98)	
**Derby 2000*	*US*	*35-74*	*431, 432, 434, 435, 436, 437*	*9*	*H*	*3811*	*P/S*	*Medical Record*[***]	*2269*	*60 (58 to 61)*	*11*
**Derby 2001*	*US*	*35-74*	*431, 432, 434, 436, 437*	*9*	*H*	*2124*	*P*	*Medical Record*[***]	*1699*	*80 (78 to 82)*	*11*
**Lakshmin. 2009*	*US*	*30-74*	*431, 432, 434, 436, 437*	*9*	*H*	*6032*	*P/S*	*Medical Record*	*3773*	*63 (61 to 64)*	*9*
			431, 432, 434, 436, 437			*4445*	*P*			*85 (84 to 86)*	
**Davenport 1996*	*UK*	*≥ 18*	*431, 433-438*	*9*	*H*	*557*	*P*	*Inpatient Register*	*529*	*95 (93 to 96)*	*11*

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
**Barer 1996*	*UK*	*-*	*431, 433, 434, 436*	*9*	*H*	*340*	*-*	*Hospital Register*	*278*	*82 (77 to 86)*	*8*
**Mant 1998*	*UK*	*-*	*431, 432.9, 434, 436, 437.0, 437.1, 437.9*	*9*	*H*	*318*	*P/S*	*Inpatient Register*	*230*	*72 (67 to 77)*	*11*
**Ives 1995*	*US*	*≥ 65*	*430, 431, 432.9, 434, 436*	*9-CM*	*H*	*79*	*-*	*Medical Record*	*71*	*90 (81 to 95)*	*8*
**Appelros 2011*	*Sweden*	*-*	*I61, I63, I64*	*10*	*D*	*98*	*-*	*Population Register*	*78*	*80 (71 to 86)*	*11*
			I61, I63, I64		*H*	*328*	*-*		*318*	*97 (94 to 98)*	
			I61, I63, I64		*H + D*	*363*	*-*		*333*	*92 (88 to 94)*	
**Koster 2013*	*Sweden*	*≥ 20*	*I61, I63, I64*	*10*	*D*	*102*	*P/S*	*Population Register*	*40*	*39 (30 to 49)*	*12*
					H	*1426*	*P/S*		*1224*	*86 (84 to 88)*	
					H + D	*1526*	*P/S*		*1264*	*83 (81 to 85)*	
**Aboa-Eboule 2013*	*France*	*-*	*I61, I63, I64, G46*	*10*	*H*	*903*	*P*	*Hospital Register*	*625*	*69 (66 to 72)*	*11*

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Studies validating codes for ischemic stroke											
Benesch 1997	US	-	433, 434, 436	9-CM	H	550	P/S	Medical Record	234	43 (38 to 47)	4
			433, 434, 436			379	P		199	53 (47 to 57)	
			434, 436			250	P/S		216	86 (82 to 90)	
			434, 436			203	P		183	90 (85 to 94)	
			433, 434			521	P/S		210	40 (36 to 45)	
			433, 434			361	P		183	51 (46 to 56)	
			433			295	P/S		18	6 (4 to 9)	
			433			176	P		16	9 (6 to 14)	
			434			226	P/S		192	85 (80 to 89)	
			434			185	P		167	91 (85 to 94)	
Goldstein 1998	US	-	434	9-CM	H	108	P	Discharge summary[***]	88	82 (73 to 88)	7
			434.x1			106	P		86	82 (73 to 87)	
			433, 434, 436			175	P		106	61 (53 to 68)	
			434, 436			127	P		104	82 (74 to 88)	
Rosamond 1999	US	45-64	433, 434, 436	9-CM	H	560	-	Medical Record	252	45 (41 to 49)	12
			434, 436			294			216	73 (68 to 78)	
			433			266			36	14 (10 to 18)	
			434			186			143	77 (70 to 82)	
			436			108			73	70 (58 to 76)	

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Rinaldi 2003	Italy	-	434, 436	9	H	180	P/S	Inpatient register	128	71 (64 to 77)	9
			434, 436			157	P		119	76 (69 to 82)	
			436			177	P/S		125	71 (64 to 77)	
			436			154	P		116	75 (68 to 81)	
Leone 2004	Italy	-	433	9	H	134	P/S	Inpatient register	8	6 (3 to 11)	11
			433			89	P		7	8 (4 to 15)	
			434			202	P/S		176	87 (82 to 91)	
			434			188	P		169	90 (85 to 93)	
			436			57	P/S		40	70 (57 to 80)	
			434, 436			259	P/S		216	83 (78 to 87)	
		-	434, 436			236	P		205	87 (82 to 91)	
			433, 434,			393	P/S		224	57 (52 to 62)	
			433, 434, 436			325	P		212	65 (60 to 70)	
Ellekjaer 1999	Norway	≥ 15	436	9	H	313	P/S	Stroke register	206	66 (61 to 71)	9
			434, 436			402			261	65 (60 to 69)	
			436			89			55	62 (51 to 71)	
Roumie 2008	US	50-84	433.x1, 434.x1, 436	9-CM	H	150	P	Medical Record	127	85 (78 to 90)	9
Johnsen 2002	Denmark	50-64	I63	10	H	113	P/S	Medical record	99	88 (80 to 92)	10
			I63, I64			313			238	76 (71 to 80)	
			I64			200			139	70 (63 to 75)	

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Krarup 2007	Denmark	-	I63, I64	10	H	138	-	Medical record	96[g]	70 (61 to 77)	9
			I64			105	-		64[g]	61 (51 to 70)	
Wright 2012	UK	-	I63	10	H	190	P/S	GP record[¶]	164	86 (81 to 90)	10
			I63, I64			309			242	78 (73 to 83)	
			I64			119			78	66 (57 to 73)	
Tirschwell 2002	US	≥ 20	433.X1, 434.X1, 436[§§]	9-CM	H+D	i) 50[¶¶]	P/S	Medical Record	45	90 (79 to 96)	9
			433.X1, 434.X1, 436[§§]			Ii) 50[¶¶]	P/S		46	91 (81 to 98)	
			433.X1, 434.X1, 436			iii) 50[¶¶]	P		44	88 (76 to 94)	
Wahl 2010	US	-	433.x1, 434.x1, 436, 437.1, 437.9	9-CM	H	200	-	Medical record	175	87 (82 to 91)	9
Haesbert 2013	France	>18	I63	10	H	329	P	Hospital register and medical record	313	95 (92 to 97)	11
**Tonolen 2007*	*Finland*	*25-74*	*433, 434, I63*	*9 +10*	*H+D*	*2711*	*P/S*	*Hospital register*	*2223*	*82 (81 to 83)*	*5*
			433, 434, 436, I63, I64			*2900*			*2407*	*83 (82 to 84)*	
Studies validating codes for haemorrhagic stroke (SAH or ICH)											
Rosamund 1999	US	45-64	430, 431	9-CM	H	63	P/S	Medical Record	46	73 (61 to 82)	12
Leone 2004	Italy	-	430, 431	9	H	152	P/S	Inpatient Register	131	86 (80 to 91)	11
Ellekjaer 1999	Norway	≥ 15	430, 431	9	H	69	P/S	Stroke Register	51	74 (62 to 83)	9

Study	Country	Age (range)	ICD code group	ICD version	Code Source	Coded events assessed (n)[†] (a)	Diagnostic position	Reference standard[‡]	Coded events confirmed (n) (b)	PPV (% & 95% CI) (b/a)	Score[§] (14)
Tonolen 2007	Finland	25-74	430, 431, I60, I61	9+10	H+D	729	P/S	Hospital Register	646	89 (86 to 91)	5
Johnsen 2002	Denmark	50-64	I60, I61	10	H	65	P/S	Medical Record	42	65 (52 to 75)	10
Studies validating codes for subarachnoid Haemorrhage (SAH)											
Tirschwell 2002	US	≥ 20	430[§§]	9-CM	H+D	i) 51[¶¶]	P/S	Medical Record	43	86 (72 to 92)	9
			430[§§]			ii) 51[¶¶]	P/S		46	89 (79 to 97)	
			430			iii)51[¶¶]	P		48	94 (84 to 98)	
Tonolen 2007	Finland	25-74	430, I60	9 + 10	H+D	253	P/S	Stroke Register	220	87 (82 to 91)	5
Kirkmann 2009	UK	-	I60	10	H	1169	P	Discharge summary	1123	96 (95 to 97)	8
Wright 2012	UK	-	I60	10	H	78	P/S	GP record[¶]	75	96 (89 to 99)	10
Studies validating codes for intracerebral haemorrhage (ICH)											
Leone 2004	Italy	-	431	9	H	110	P/S	Inpatient register	82	75 (66 to 82)	11
						102	P		78	76 (67 to 84)	
Ellekjaer 1999	Norway	≥ 15	431	9	H	56	P/S	Stroke Register	40	71 (59 to 82)	9
Tonolen 2007	Finland	25-74	431, I61	9 + 10	H + D	476	P/S	Stroke Register	413	87 (83 to 90)	5
Kirkmann 2009	UK	-	I61	10	H	978	P	Discharge summary	938	96 (94 to 97)	8

PPV: Positive Predictive Value; H: Hospital data; D: Death certificates; H+D: both; P: Primary position code; P/S: Primary or Secondary position code.

*_Italics_ indicate studies using miscellaneous groups of codes (not included in either 'stroke codes' or 'all cerebrovascular diseases codes' groups, Figure 2.1), or studies excluding SAH

[†]Number of ICD coded events compared against the reference standard.

[‡]Population register: population based stroke register, Hospital register: inpatient and outpatient stroke register, Inpatient register: inpatient stroke register. Medical Record: definite or probable stroke diagnoses confirmed by review of medical records (excludes 'possible' stroke). Medical records were reviewed by stroke physicians or neurologists, unless otherwise specified.

[§]Quality score (total 14). See Appendix 2.7.12 for questions and scoring methods.

[¶]Medical record reviewed by 'cardiovascular researchers'.

[**]Medical record reviewed by 'trained data abstractors'.

[††]431, 432, 434, 436 primary position, or rehabilitation code primary position AND 430-438 secondary position, OR 433, 434 primary position AND 430-438 secondary position.

[‡‡]Mean value calculated from published data.

[§§]If > 1 code per discharge they were chosen in the following hierarchy: SAH>ICH>IS>TIA

[¶¶]One code for each discharge chosen from: i) First 9 discharge diagnoses ii) First 2 discharge diagnoses iii) Primary discharge code chosen

[***]Abstracts of the medical record reviewed by 'study physician'.

2.4.3 Quality assessment

Results of the quality assessment are displayed in Table 2.2. Quality scores ranged from 4 to 12 (median 9, interquartile range 8 to 11). With respect to reporting quality, participant selection criteria and coding algorithms were generally well reported, but only ten studies acknowledged the potential for uncertainty of the reference standard diagnosis in their results.(Rosamond et al. 1999, Derby et al. 2001, Mayo et al. 1993, Klatsky et al. 2005, Wahl et al. 2010, Appelros and Terent 2011, Sinha et al. 2008, Davenport et al. 1996, Mant et al. 1997, Koster et al. 2013b) With respect to generalisability to the UK population, only eight studies were conducted in the UK. However, all the other studies were based in high income countries, among populations of predominantly European origin with broadly similar health care provision, and are therefore likely to be broadly generalizable (from a global perspective) to population-based studies in these types of settings (including the UK). Of the UK-based studies, two had suboptimal generalisability because all coded discharges were taken from a single hospital department,(Panayiotou et al. 1993, Hasan et al. 1995) while for the other six generalisability was unclear due to incomplete reporting.(Sinha et al. 2008, Wright et al. 2012, Davenport et al. 1996, Mant et al. 1997, Barer 1996, Kirkman et al. 2009)

With respect to risk of bias, only five studies achieved the optimum score.(Rosamond et al. 1999, Appelros and Terent 2011, Stegmayr and Asplund 1992, Aboa-Eboule et al. 2013, Harriss et al. 2011) Incomplete reference standard data (due to a variable proportion of missing or irretrievable records)(Ives et al. 1995, Lakshminarayan et al. 2009, Leibson et al. 1994, Roumie et al. 2008, Derby et al. 2001, Liu et al. 1999, Klatsky et al. 2005, Goldstein 1998, Benesch et al. 1997, Johnsen et al. 2002, Krarup et al. 2007, Tolonen et al. 2007, Ellekjær et al. 1999, Spolaore et al. 2005, Rinaldi et al. 2003, Palmieri et al. 2007b, Sinha et al. 2008, Wright et al. 2012, Barer 1996, Kirkman et al. 2009, Koster et al. 2013b) and lack of or inadequate blinding of adjudicators to the coded diagnosis (Ives et al. 1995, Lakshminarayan et al. 2009, Reker et al. 2001, Roumie et al. 2008, Derby et al. 2001, Liu et al. 1999, Mayo et al. 1993, Klatsky et al. 2005, Goldstein 1998, Benesch et al. 1997, Johnsen et al. 2002, Appelros and Terent 2011, Tolonen et al. 2007, Ellekjær et al. 1999, Leone et al. 2004, Spolaore et al. 2005, Rinaldi et al. 2003, Haesebaert et

al. 2013, Sinha et al. 2008, Wright et al. 2012, Panayiotou et al. 1993, Hasan et al. 1995) were the most common potential causes of bias.

Table 2.2 Quality assessment of included studies

Study	Reporting						Generalisability				Risk of bias							Total score
	Selection Criteria*	Index test†	Reference standard‡	Participants excluded§	Unclear results¶	Total (5)	Country**	Population selected††	Reference standard‡‡	Total (3)	Selection bias§§	Blind¶¶	Independ.***	Differential verification†††	Reference standard‡‡‡	Timing§§§	Total (6)	(14)
Reker	☺	☺	☹	☺	☹	3	☹	?	?	0	☹	☹	☺	☺	?	☺	3	6
Aboa-Eboule	☺	☺	☺	☺	☹	4	☹	?	☺	1	☺	☺	☺	☺	☺	☺	6	11
Sinha	☺	☺	☺	☺	☺	5	☺	?	☺	2	☹	☹	☺	☺	☺	☺	4	11
Wahl	☺	☹	☹	☺	☺	3	☹	☺	?	1	☺	☺	☺	☺	?	☺	5	9
Rinaldi	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☹	☹	☺	☹	☺	☺	3	9
Benesch	☹	☺	☺	☹	☹	2	☹	?	?	0	?	☹	☺	☺	?	☹	2	4
Klatsky	☺	☺	☺	☺	☺	5	☹	?	☺	1	☹	☹	☺	☺	☺	☺	4	10
Goldstein	☺	☺	☹	☺	☹	3	☹	?	?	0	☹	☹	☺	☺	☺	☺	4	7
Koster	☺	☺	☺	☺	☺	5	☹	☺	☺	2	☹	☺	☺	☺	☺	☺	5	12
Hasan	☺	☺	☺	☺	☹	4	☺	☹	?	1	☺	☹	☺	☺	?	☺	4	9
Derby	☺	☺	☺	☺	☺	5	☹	☺	☺	2	☹	☹	☺	☺	☺	☺	4	11
Derby	☺	☺	☺	☺	☺	5	☹	☺	☺	2	☹	☹	☺	☺	☺	☺	4	11
Rosamond	☺	☹	☺	☺	☺	4	☹	☺	☺	2	☺	☺	☺	☺	☺	☺	6	12
Barer	☺	☹	☹	☹	☹	1	☺	?	☺	2	?	☺	☺	☺	☺	☺	5	8
Haesebart	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☺	☹	☺	☺	☺	☺	5	11
Johnsen	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☹	☹	☺	☺	☺	☺	4	10
Appelros	☺	☹	☺	☺	☹	3	☹	☺	☺	2	☺	☺	☺	☺	☺	☺	6	11
Tirschwell	☺	☺	☹	☺	☹	3	☹	☺	?	1	☺	☺	☺	☺	?	☺	5	9
Davenport	☺	☺	☺	☺	☺	5	☺	?	?	1	☺	☺	☺	☹	☺	☺	5	11
Stegmayr	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☺	☺	☺	☺	☺	☺	6	12

Liu	☺	☺	☺	☹	☹	3	☹	?	☺	1	?	☹	☺	☺	☺	☺	4	8
Kirkman	☺	☺	☺	☹	☹	3	☺	☺	☹	2	?	☹	☺	☺	☹	☺	3	8
Palmieri	☺	☺	☺	☹	☹	3	☹	☺	?	1	?	☹	☺	☺	☺	☺	4	8
Mant	☺	☺	☺	☺	☹	4	☺	?	☺	2	☺	☺	☺	☹	☺	☺	5	11
Mayo	☺	☺	☹	☺	☺	4	☹	?	?	0	☺	☹	☺	☺	?	☺	4	8
Panayiotou	☺	☺	☹	☺	☹	3	☺	☹	☺	2	☹	☹	☺	☺	?	☺	3	8
Wright	☺	☺	☺	☺	☺	5	☺	☺	?	2	☹	☹	☺	☺	?	☺	3	10
Sporalore	☺	☺	☺	☹	☹	3	☹	☺	☺	2	?	☹	☺	☺	☺	☺	4	9
Roumie	☺	☺	☺	☺	☹	4	☹	?	☺	1	☹	☹	☺	☺	☺	☺	4	9
Krarup	☺	☹	☺	☺	☹	3	☹	☺	☺	2	☹	☺	☺	☺	?	☺	4	9
Harriss	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☺	☺	☺	☺	☺	☺	6	12
Ellekjaer	☺	☺	☺	☹	☹	3	☹	☺	☺	2	?	☹	☺	☺	☺	☺	4	9
Tonolen	☺	☺	☹	☹	☹	2	☹	☺	?	1	?	☹	☺	?	?	☺	2	5
Leone	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☺	☹	☺	☺	☺	☺	5	11
Ives	☺	☹	☺	☹	☹	2	☹	☺	☺	2	?	☹	☺	☺	☺	☺	4	8
Lakshminar.	☺	☺	☺	☹	☹	3	☹	☺	☺	2	?	☹	☺	☺	☺	☺	4	9
Leibson	☺	☺	☺	☺	☹	4	☹	☺	☺	2	☺	☺	☺	☹	☺	☺	5	11

☺High reporting quality/High generalizability/Low risk of bias ☹ Low reporting quality/Low generalizability/High risk of bias ? Unclear generalizability/Unclear risk of bias. Rules for assessment of study quality are displayed in Appendix 2.7.12.

*Were the selection criteria clearly described?

†Was execution of the index test described in sufficient detail to allow replication of the test?

‡Was execution of the reference standard described in sufficient detail to permit its replication?

§If participants were excluded from the final analysis, were they described and were the reasons for this exclusion explained?

¶Were uninterpretable/intermediate results reported?

**Was the study UK-based?

[††]Was the spectrum of patients selected representative of the patients who will receive the diagnosis in practice?

[‡‡]Were the same clinical data available when test results were interpreted as would be available when used in practice?

[§§]Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?

[¶¶]Were the reference standard results interpreted without knowledge of the results of the index test?

[***]Was the reference standard independent of the index test?

[†††]Did all patients receive the same reference standard regardless of the index test result?

[‡‡‡] Is the reference standard likely to correctly classify the target condition?

[§§§]Was the time period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests.

2.4.4 Accuracy of ICD coded events

The range of PPVs reported for various codes used to identify stroke or one of its main pathological types was very broad, reflecting considerable heterogeneity of study characteristics. Results were particularly variable for all stroke (PPV 31-97%) and for ischaemic stroke (PPV 6-95%), while they appeared more consistent for haemorrhagic stroke (PPV 73-89%), SAH (PPV 86-96%) and ICH (PPV 71-96%), although based on fewer studies.

Within study comparisons. Only six studies used a population based reference standard and, of these, only four (all from Scandinavian countries)(Appelros, Hogeras and Terent 2003, Ellekjær et al. 1999, Stegmayr and Asplund 1992, Koster et al. 2013a) provided sufficient data to calculate sensitivity, specificity and negative predictive value (NPV) of codes for stroke. Sensitivities for identifying stroke were around 80% or more using general cerebrovascular or stroke specific codes from either hospital data or hospital data combined with death certificates, but – unsurprisingly – sensitivity was much lower for death certificates alone (Table 2.3). There were no data on sensitivity for the main pathological types of stroke.

Where calculable, specificity and NPV were uniformly high (range 96-99.9%), reflecting the relatively small proportion of false negative strokes (amongst all non stroke and code negative numbers, respectively).(Appelros and Terent 2011, Ellekjær et al. 1999, Stegmayr and Asplund 1992, Koster et al. 2013a)

Table 2.3 Sensitivity, Specificity, PPV, and NPV of codes for stroke versus a population reference standard

Study	Country	ICD code group	Code source	Reference standard*		Coded events compared against the reference standard† (n)				Sensitivity (% & 95% CI)	Specificity (% & 95% CI)	PPV (% & 95% CI)	NPV (% & 95% CI)
				Stroke	*No stroke*	*TP*	*FP*	*TN*	*FN*				
Appelros	Sweden	I61, I63, I64	D	377	123,126	78	20	123,106	299	21 (71-25)	99.9 (99.97-99.98)	80 (71-86)	99.7 (99.75-99.78)
			H			318	10	123,116	59	84 (80-88)	99.9 (99.98-99.99)	97 (94-98)	99.9 (99.93-99.96)
			H + D			333	30	123,096	44	88 (85-91)	99.9 (99.96-99.98)	92 (88-94)	99.9 (99.95-99.97)
Koster	Sweden	I61, I63, I64	D	1,351	508,649	40	62	508,587	1,311	3 (2-4)	99.9 (99.98-99.99)	39 (30-49)	99.7 (99.73-99.76)
			H			1224	202	508,447	127	91 (89-92)	99.9 (99.95-99.97)	86 (84-88)	99.9 (99.97-99.98)
			H + D			1264	264	508,385	87	94 (92-95)	99.9 (99.94-99.95)	83 (81-85)	99.9 (99.98-99.99)
Stegmayr	Sweden	430-438	D	4528	94,556	812	87	94,469	3,716	18 (17-19)	99.9 (99.88-99.92)	90 (88-92)	96.2 (96.09-96.33)
			H			3492	1,609	92,947	1,036	78 (76-78)	98.3 (98.21-98.38)	69 (67-70)	98.9 (98.83-98.96)
Ellekjaer	Norway	430-438	H	430	69,570	369	390	69,180	61	86 (82-89)	99.4 (99.38-99.49)	49 (45-52)	99.9(99.88-99.93)
		430,431,434,436	H			347	161	69,409	83	81 (77-84)	99.7 (99.73-99.80)	68 (64-72)	99.9 (99.85-99.90)

D: death certificates; H: hospital data; H + D: both

*Population stroke register †TP=True Positive cases; FP=False Positive cases; TN=True Negative cases; FN= False Negative cases.

Several within study comparisons showed that the groups of codes with the highest PPVs (68-90%) for all types of stroke combined were 430, 431, 434, 436 (ICD-9) or I60, I61, I63, I64 (ICD-10) (Table 2.4). Compared with general cerebrovascular codes (ICD-9 430-438, or ICD-10 I60-I69+/-G45), selection of these stroke specific codes gave consistently higher PPVs (absolute increase of 17-30%) (Table 2.4). Stroke specific codes inevitably identified fewer coded events than general cerebrovascular ones (numbers fell by a third to over a half, Table 2.4), but the impact on sensitivity appeared limited (absolute decrease of 5%) in the one study that provided data on this.(Ellekjær et al. 1999)

Table 2.4 Effect on PPV of codes used to identify stroke: within-study comparisons*

Study	Codes	Diagnosis sought	Coded events	PPV (%, & 95% CI)
Johnsen 2002	I60-I69 + G45	CVD	565	58 (58-62)
	I60, I61, I63, I64	Stroke	378	79 (75-83)
Krarup 2007†	I60-I69 + G45	CVD	236	69 (59-71)
	I60, I61, I63, I64	Stroke	164	86 (76-88)
Ellekjaer 1999	430-438	CVD	759	49 (45-52)
	430, 431, 434, 436	Stroke	508	68 (64-72)
Leone 2004	430-438	CVD	1017	60 (57-63)
	430, 431, 434, 436	Stroke	411	90 (87-93)

CVD= Cerebrovascular disease

*If there was more than one result per code group, results are shown for the largest number of cases assessed.

†Mean PPV taken from range of values in original publication.

For identifying ischaemic stroke, codes I63 (ICD-10) or 434 (ICD-9) achieved reasonably high PPVs (range 66 to 88%), (Rosamond et al. 1999, Goldstein 1998, Benesch et al. 1997, Johnsen et al. 2002, Leone et al. 2004, Haesebaert et al. 2013, Wright et al. 2012) while code 433 (ICD-9) performed consistently poorly in studies which assessed it (PPV range 6% to 14%).(Rosamond et al. 1999, Benesch et al. 1997, Leone et al. 2004) The addition of codes for unspecified type of stroke (436 [ICD-9] or I64 [ICD-10]) to ischaemic stroke codes increased the number of coded events identified within each study, with in general either no change or only a few % absolute decrease in PPV (Table 2.5).(Rosamond et al. 1999, Goldstein 1998,

Benesch et al. 1997, Johnsen et al. 2002, Krarup et al. 2007, Tolonen et al. 2007, Ellekjær et al. 1999, Leone et al. 2004, Rinaldi et al. 2003, Wright et al. 2012)

Table 2.5 Effect on PPV of codes used to identify ischaemic stroke: within study comparisons*

Study	Codes	Diagnosis sought	Coded events	PPV (%, & 95% CI)
Johnsen 2002	I63	Ischaemic stroke	113	88 (80-93)
	I64	Unspecified stroke	200	70 (63-76)
	I63, I64	Ischaemic and unspecified stroke	313	76 (71-80)
Wright 2012	I63	Ischaemic stroke	190	86 (81-91)
	I64	Unspecified stroke	119	66 (57-73)
	I63, I64	Ischaemic and unspecified stroke	309	78 (73-83)
Ellekjaer 1999	434	Ischaemic stroke	313	66 (60-71)
	436	Unspecified stroke	89	62 (51-71)
	434, 436	Ischaemic and unspecified stroke	402	65 (60-69)
Leone 2004	434	Ischaemic stroke	202	87 (82-91)
	433	Ischaemic stroke	134	6 (3-11)
	436	Unspecified stroke	57	70 (57-80)
	434, 436	Ischaemic and unspecified stroke	259	83 (78-87)
	433, 434, 436	Ischaemic and unspecified stroke	393	57 (52-62)
Rosamond 1999	434	Ischaemic stroke	186	77 (70-82)
	433	Ischaemic stroke	266	14 (10-18)
	436	Unspecified stroke	108	70 (52-76)
	434, 436	Ischaemic and unspecified stroke	294	73 (68-78)
	433, 434, 436	Ischaemic and unspecified stroke	560	45 (41-49)
Benesch 1997	434	Ischaemic stroke	226	85 (79-89)
	433	Ischaemic stroke	295	6 (4-9)
	434, 436	Ischaemic and unspecified stroke	250	86 (82-90)
	433, 434, 436	Ischaemic and unspecified stroke	550	43 (38-47)
Krarup 2007†	I64	Unspecified stroke	105	60 (50-69)
	I63, I64	Ischaemic and unspecified stroke	138	70 (61-77)
Rinaldi 2003	436	Unspecified stroke	177	71 (64-77)
	434, 436	Ischaemic and unspecified stroke	180	71 (64-77)
Tonolen 2007	433, 434, I63	Ischaemic stroke	2711	82 (81-83)
	433, 434, 436, I63, I64	Ischaemic and unspecified stroke	2900	83 (82-84)
Goldstein 1998	434	Ischaemic stroke	108	82 (74-88)
	434.x1	Ischaemic stroke	106	82 (74-88)
	434, 436	Ischaemic and unspecified stroke	127	82 (74-88)
	433, 434, 436	Ischaemic and unspecified stroke	175	61 (53-68)

*If there was more than one result per code, results are shown for the largest number of cases assessed.

†Mean PPV taken from range of values in original publication.

Eight studies (all of ICD-9 codes) assessed influence of coding position on PPV for a variety of ICD-9 code groups (cerebrovascular disease codes, ischaemic stroke codes, or haemorrhagic stroke codes).(Lakshminarayan et al. 2009, Leibson et al. 1994, Roumie et al. 2008, Liu et al. 1999, Tirschwell and Longstreth 2002, Benesch et al. 1997, Leone et al. 2004, Rinaldi et al. 2003) Restriction to the primary position code (versus inclusion of codes from the primary or secondary diagnostic position) increased the PPV, but by no more than about 5-10% in all but two studies (Lakshminarayan et al. 2009, Liu et al. 1999) (Table 2.6).

Table 2.6 Influence of diagnostic position on PPV

Study	ICD code group	Diagnostic position	Coded events (n)	PPV (% & 95% CI)
Leone 2004	430-438	Primary or Secondary	1017	60 (57-63)
		Primary	833	66 (63-69)
Lakshminarayan 2009	431, 432, 434, 436, 437	Primary or Secondary	6032	63 (61-64)
		Primary	4445	85 (84-86)
Leibson 1994	430-438	Primary or Secondary	462	54 (49-58)
		Primary	377	60 (55-64)
Roumie 2008	430, 431, 433.x1, 434.x1, 436	Primary or Secondary	231	85 (84-92)
		Primary	203	97 (93-98)
Liu 1999	430-438	Primary or Secondary	862	56 (53-60)
		Primary	621	87 (63-70)
Rinaldi 2003	434, 436	Primary or Secondary	180	71 (64-77)
		Primary	157	76 (69-82)
Tirschwell 2002	430	Primary or Secondary*	51*	86 (74-93)
		Primary*	51*	94 (84-98)
Benesch 1997	433, 434, 436	Primary or Secondary	550	43 (39-47)
		Primary	379	53 (48-58)

*A single code was selected from 51 hospital discharges: Primary or Secondary: used the hierarchy SAH>ICH>IS>TIA to select a single code for each discharge; Primary: selected the primary discharge code.

It was not possible directly to assess the influence of code position on sensitivity, but restriction to the primary position reduced the number of coded events identified by around 10-30%.

Comparisons between groups of studies reporting PPV for stroke and its main types. The PPV of codes for stroke and its main types, stratified according to the code group(s) selected (see below), are displayed in Figures 2.3 to 2.5. They display results of studies which identified: stroke events using either a broad selection of cerebrovascular codes or stroke-specific codes (Figure 2.3); ischaemic stroke events, using either codes for ischaemic and unspecified type of stroke or for ischaemic stroke alone (Figure 2.4.); and haemorrhagic stroke events using codes for ICH and SAH together or separately (Figure 2.5.). Informed by our within-study comparisons, results exclude studies which included the poorly performing ICD-9 code 433 among the stroke-specific or ischaemic stroke codes, except those which used the clinical modification 433.x1 (Figure 2.1, Table 2.5, Figure 2.4).

For each of stroke and its main pathological types, PPVs of >90% were achieved in some studies (Figures 2.3 to 2.5). In line with results from within-study comparisons (Table 2.4), stroke-specific codes yielded higher PPVs for stroke (range 68-90%) than general cerebrovascular disease codes (range 31-80%) (Figure 2.3), while PPVs for ischaemic stroke were slightly higher with codes for ischaemic stroke alone (range 66-95%) than with codes for ischaemic and unspecified stroke (range 65-90%), but identified smaller numbers of outcomes (Figure 2.4). Codes for haemorrhagic stroke, and for ICH and SAH separately, performed consistently well or very well (PPV range 65-96%) (Figure 2.5). In general, ICD-10 appeared to perform better than ICD-9 codes, except where the ‘clinical modification’ (ICD-9-CM, see Figure 2.1) was available. Studies from the UK, yielding data that might be considered most informative for UK Biobank, reported PPVs of 78% and 86% for ischaemic stroke in one study (Wright et al. 2012) (the lower value when codes for unspecified stroke were included), 96% for SAH in two studies (Wright et al. 2012, Kirkman et al. 2009) and 96% for ICH in one study.(Kirkman et al. 2009) The quality scores did not appear to influence PPV (Figure 2.3 to Figure 2.5).

First author	Code version	Country	Year	Code source	Diagnostic position	Quality (14)	x/y	PPV (95% CI)
Cerebrovascular disease codes*								
Mayo†‡	9	Canada	1993	H	P	8	73/96	0.76 (0.67, 0.83)
Liu	9	Canada	1999	H	P/S	8	487/862	0.56 (0.53, 0.60)
Leibson	9	US	1994	H	P/S	11	249/462	0.54 (0.49, 0.58)
Rosamond	9	US	1999	H	-	12	326/1058	0.31 (0.28, 0.34)
Reker	9	US	2001	H	-	6	279/671	0.42 (0.38, 0.45)
Klatsky	9	US	2005	H	P	10	2494/3239	0.77 (0.76, 0.78)
Leone	9	Italy	2004	H	P/S	11	609/1017	0.60 (0.57, 0.63)
Sporalore§	9	Italy	2005	H	P/S	9	1296/3619	0.36 (0.34, 0.37)
Palmieri†¶	9	Italy	2007	H+D	P/S	8	1173/2793	0.42 (0.40, 0.44)
Stegmayr†	9	Sweden	1992	H+D	P/S	12	4304/6000	0.72 (0.71, 0.73)
Ellekjaer	9	Norway	1999	H	P/S	9	369/759	0.49 (0.45, 0.52)
Panayioutou	9	UK	1993	H	P	8	94/117	0.80 (0.72, 0.87)
Hasan	9‡	UK	1996	H	P	9	113/166	0.68 (0.61, 0.75)
Harris	9+10	Australia	2010	D	P	12	72/119	0.61 (0.52, 0.69)
Johnsen	10	Denmark	2002	H	P/S	10	325/565	0.58 (0.53, 0.62)
Krarup†	10	Denmark	2007	H	-	9	153/236	0.65 (0.59, 0.71)
Sihna	10	UK	2008	H+D	P/S	11	191/250	0.76 (0.71, 0.81)
Stroke-specific codes#								
Leone	9	Italy	2004	H	P/S	11	371/411	0.90 (0.87, 0.93)
Ellekjaer	9	Norway	1999	H	P/S	9	347/508	0.68 (0.64, 0.72)
Roumie	9-CM	US	2008	H	P/S	9	205/231	0.89 (0.84, 0.92)
Johnsen	10	Denmark	2002	H	P/S	10	299/378	0.79 (0.75, 0.83)
Krarup†	10	Denmark	2007	H	-	9	136/164	0.83 (0.76, 0.88)

Figure 2.3 Positive predictive value of codes for stroke.

Figure 2.3 Positive predictive value of codes for stroke

H: hospital data, D: death certificates, H+D: hospital data and death certificates; x =number of coded events confirmed as 'true cases' by the reference standard; y= total number of coded events; x/y = PPV.

Circles represent PPVs, and horizontal lines denote 95% confidence intervals (CIs). Circle size is proportional to the inverse variance of the PPV. Where more than one result was available for a particular study, the result for the largest number of coded events validated is shown.

[*]Cerebrovascular codes: I60-I69+/-G45 (ICD-10) or 430-438 (ICD-9), unless otherwise specified

[†] Mean PPV (taken from the range published in the study)

[‡]Excluding codes 435 (TIA) and 438 (sequelae of cerebrovascular disease)

[§]Excluding code 435 (TIA) and including code 342 (hemiplegia and hemiparesis)

[¶]Excluding code 435 (TIA)

[#]Stroke-specific: 160, 161, 163, 164 (ICD-10), 430, 431, 434, 436 (ICD-9), 430, 431, 433.x1, 434.x1 (ICD-9-CM)

[¥]Ischaemic stoke and unspecified stroke: I63, I64 (ICD-10), 434, 436 (ICD-9), 433.x1, 434.x1, 436 (ICD-9-CM)

[**]Ischaemic stroke:163 (ICD-10), 434 (ICD-9), 433.x1, 434.x1 (ICD-9-CM)

[††]Haemorrhagic stroke:I60, I61 (ICD-10), 430, 431 (ICD-9)

[‡‡]Subarachnoid haemorrhage stroke:I60 (ICD-10), 430 (ICD-9)

[¶¶]Intracerebral haemorrhage stroke codes:I61 (ICD-10), 431 (ICD-9)

First author	Code version	Country	Year	Code source	Diagnostic position	Quality (14)	x/y	PPV (95% CI)
Ischaemic stroke and unspecified stroke codes**[¥]								
Benesch	9	US	1997	H	P/S	4	216/250	0.86 (0.82, 0.90)
Goldstein	9	US	1998	H	P	7	104/127	0.82 (0.74, 0.88)
Rosamond	9	US	1999	H	-	12	216/294	0.73 (0.68, 0.78)
Rinaldi	9	Italy	2003	H	P/S	9	128/180	0.71 (0.64, 0.77)
Leone	9	Italy	2004	H	P/S	11	216/259	0.83 (0.78, 0.87)
Ellekjaer	9	Norway	1999	H	P/S	9	261/402	0.65 (0.60, 0.69)
Tirschwell	9-CM	US	2002	H+D	P/S	9	45/50	0.90 (0.79, 0.96)
Roumie	9-CM	US	2008	H	P	9	127/150	0.85 (0.78, 0.90)
Johnsen	10	Denmark	2002	H	P/S	10	238/313	0.76 (0.71, 0.80)
Krarup†	10	Denmark	2007	H	-	9	96/138	0.70 (0.61, 0.77)
Wright	10	UK	2012	H	P/S	10	242/309	0.78 (0.73, 0.83)
Ischaemic stroke codes**[**]								
Benesch	9	US	1997	H	P/S	4	192/226	0.85 (0.80, 0.89)
Goldstein	9	US	1998	H	P	7	88/108	0.82 (0.74, 0.88)
Rosamond	9	US	1999	H	-	12	143/186	0.77 (0.70, 0.82)
Leone	9	Italy	2004	H	P/S	11	176/202	0.87 (0.82, 0.91)
Ellekjaer	9	Norway	1999	H	P/S	9	206/313	0.66 (0.60, 0.71)
Wahl	9-CM	US	2010	H	-	9	150/156	0.95 (0.91, 0.98)
Johnsen	10	Denmark	2002	H	P/S	10	99/113	0.88 (0.80, 0.93)
Haesebert	10	France	2013	H	P	11	313/329	0.95 (0.92, 0.97)
Wright	10	UK	2012	H	P/S	10	164/190	0.86 (0.81, 0.90)

Figure 2.4 Positive predictive values of codes for ischaemic stroke

Figure 2.4 Positive predictive value of codes for ischaemic stroke

H: hospital data, D: death certificates, H+D: hospital data and death certificates; x =number of coded events confirmed as 'true cases' by the reference standard; y= total number of coded events; x/y = PPV.

Circles represent PPVs, and horizontal lines denote 95% confidence intervals (CIs).

Circle size is proportional to the inverse variance of the PPV.

Where more than one result was available for a particular study, the result for the largest number of coded events validated is shown.

[†]Mean PPV (taken from the range published in the study

[¥]Ischaemic stoke and unspecified stroke codes: I63, I64 (ICD-10), 434, 436 (ICD-9), 433.x1, 434.x1, 436 (ICD-9-CM)

[**]Ischaemic stroke codes:163 (ICD-10), 434 (ICD-9), 433.x1, 434.x1 (ICD-9-CM)

First author	Code version	Country	Year	Code source	Diagnostic position	Quality (14)	x/y	PPV (95% CI)
Haemorrhagic stroke (ICH or SAH) codes††								
Rosamond	9	US	1999	H	P/S	12	46/63	0.73 (0.61, 0.82)
Leone	9	Italy	2004	H	P/S	11	131/152	0.86 (0.80, 0.91)
Ellekjaer	9	Norway	1999	H	P/S	9	51/69	0.74 (0.63, 0.83)
Tonolen	9+10	Finland	2007	H+D	P/S	5	646/729	0.89 (0.86, 0.91)
Johnsen	10	Denmark	2002	H	P/S	10	42/65	0.65 (0.53, 0.75)
Subarachnoid haemorrhage stroke codes‡‡								
Tirschwell	9	US	2002	H	P/S	9	44/51	0.86 (0.74, 0.93)
Tonolen	9+10	Finland	2007	H+D	P/S	5	220/253	0.87 (0.82, 0.91)
Kirkman	10	UK	2009	H	P	8	1123/1169	0.96 (0.95, 0.97)
Wright	10	UK	2012	H	P/S	10	75/78	0.96 (0.89, 0.99)
Intracerebral haemorrhage stroke codes¶								
Leone	9	Italy	2004	H	P/S	11	82/110	0.75 (0.66, 0.82)
Ellekjaer	9	Norway	1999	H	P/S	9	40/56	0.71 (0.59, 0.82)
Tonolen	9+10	Finland	2007	H+D	P/S	5	413/476	0.87 (0.83, 0.90)
Kirkman	10	UK	2009	H	P	8	938/978	0.96 (0.94, 0.97)

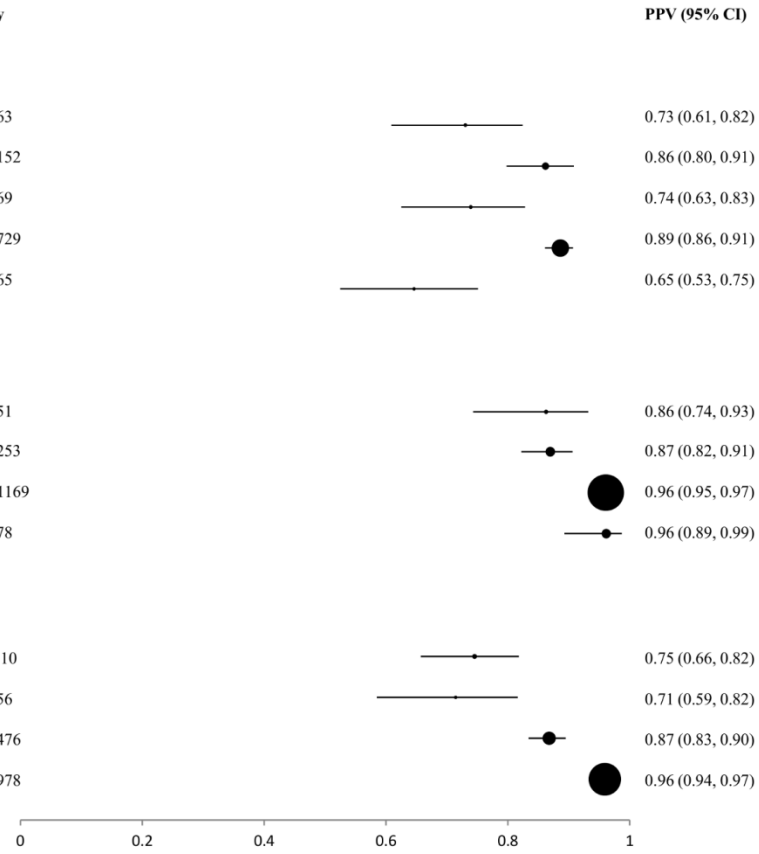

Figure 2.5 Positive predictive values of codes for haemorrhagic stroke

Figure 2.6 Positive predictive values of codes for haemorrhagic stroke

H: hospital data, D: death certificates, H+D: hospital data and death certificates;

x =number of coded events confirmed as 'true cases' by the reference standard; y= total number of coded events; x/y = PPV.

Circles represent PPVs, and horizontal lines denote 95% confidence intervals (CIs).

Circle size is proportional to the inverse variance of the PPV.

Where more than one result was available for a particular study, the result for the largest number of coded events validated is shown.

[†]Mean PPV (taken from the range published in the study)

[††]Haemorrhagic stroke codes:I60, I61 (ICD-10), 430, 431 (ICD-9)

[‡‡]Subarachnoid haemorrhage stroke codes:I60 (ICD-10), 430 (ICD-9)

[¶¶]Intracerebral haemorrhage stroke codes:I61 (ICD-10), 431 (ICD-9)

Selection of the best code using a code hierarchy. Two studies used a 'code hierarchy' to select a single stroke code when more than one was used for an individual hospital admission.(Roumie et al. 2008, Tirschwell and Longstreth 2002) These studies selected the single 'best code' for each case, based on presumed coding accuracy (SAH>ICH>ischaemic stroke>transient ischaemic attack [TIA]). This approach was no more accurate than selection of the primary position code in one study,(Tirschwell and Longstreth 2002) and less accurate than selection of the primary position code in another.(Roumie et al. 2008)(Table 2.6)

Distinguishing ischaemic stroke subtypes. Very few studies assessed accuracy of ICD codes for more detailed ischaemic stroke subtypes, and none assessed accuracy for subtypes of SAH or ICH. One study found that out of 106 coded events for ischaemic stroke subtypes, >70% had unspecified ischaemic stroke subtype codes.(Goldstein 1998) The PPV of the cardiac embolism subtype code was 73% (based on only 11 coded events), but PPVs for other ischaemic subtypes were not reported. Another study attempted to classify ischaemic strokes into four subtypes (lacunar stroke, cardiac embolism, large artery atherosclerosis and other) based on the hospital discharge abstract (which was used to generate the ICD codes) rather than the codes themselves.(Aboa-Eboule et al. 2013) This approach produced PPVs of 66-87% (highest for cardiac embolism and lacunar ischaemic stroke), and sensitivities of 67-74% (highest for cardiac embolism and large artery atherosclerosis).

2.4.5 Studies of Read-coded primary care data.

Two UK-based studies reported PPVs of Read codes from primary care data, one for ischaemic and one for haemorrhagic stroke (Table 2.7). Neither study reported code sensitivity. (Ruigomez et al. 2010, Gaist et al. 2013) PPV was 89% for ischaemic stroke and 82% for haemorrhagic stroke, increasing to 90% for haemorrhagic stroke with exclusion of haemorrhagic codes which overlapped with antithrombotic drug prescription codes.

Table 2.7 Included primary care Read-coded studies: characteristics and results

Study	Country	Age (range)	Read code(s) assessed	Coded events assessed (n)	Diagnosis sought	Coded events confirmed (n)	Reference standard	PPV (% & 95% CI)
Rugiomez 2010*	UK	40-84	G66..00 to G669.00[†]	120	Ischaemic stroke	93	GP questionnaire/	78 (69 to 85)
			G64..00 to G64z.00[‡]	28	Ischaemic stroke	25	GP record	89 (72 to 98)
Gaist 2013[§]	UK	20-89	41 codes[¶]	306	ICH & SAH	251	GP questionnaire/	82 (77 to 86)
				156	SAH	142	GP record	91 (85 to 95)
				150	ICH	109		73 (65 to 80)

[*]This study excluded individuals who were not admitted to hospital, who had cancer or a previous cerebrovascular event, or who also had Read codes for haemorrhagic stroke.

[†]Read codes for unspecified stroke.

[‡]Read codes for ischaemic stroke.

[§]This study excluded Read code events if the primary care record 'free text' suggested that events were; ischaemic; secondary to trauma; subdural haemorrhage; prevalent rather than incident; due to cancer, or events occurred in hospital.

[¶]Diagnostic codes for Subarachnoid haemorrhage (SAH), Intracerebral haemorrhage (ICH), unspecified haemorrhage and sequelae, Procedure codes for evacuation/aspiration of haematoma, and Process of care codes for 'History of subarachnoid haemorrhage.' Gaist et.al, Pharmacoepidemiology and Drug Safety 2013;22:176-182:Appendix I (online supplement).

2.4.6 Combining multiple data sources

None of the included studies assessed the combination of primary care codes with hospital or death certificate codes for stroke or its main types. A few excluded studies compared primary care and hospital codes to search for stroke plus TIA.(Mant et al. 2003, Tu et al. 2013) A UK study found that, compared to hospital ICD codes for stroke plus TIA in a primary care population of ~5800 individuals, primary care Read codes increased sensitivity and decreased PPV by absolute values of 53% and 17% respectively.(Mant et al. 2003) Similarly, a community-based study in Canada found that combining primary care physician billing data with hospital ICD codes detected more stroke/TIA events, but with lower PPV, compared to ICD codes alone: sensitivity for combined data sources was 78% (95% CI 66%-83%) versus 37% (95% CI 28%-46%) for ICD codes alone; PPV for combined data sources was 40% (95% CI 33%-46%) versus 81% (95% CI 70%-92%) for ICD codes alone.(Tu et al. 2013)

Two UK studies explored the possibility of using medical record extracts to reduce the proportion of unspecified stroke codes (I64).(Wright et al. 2012, Flynn et al. 2010) In one, the primary care record held information to classify 74% of ICD-coded 'unspecified strokes' as ischaemic or haemorrhagic.(Wright et al. 2012) In the other, CT brain scan reports were used to assign ~ 8400 stroke cases (identified by ischaemic stroke, intracerebral haemorrhage or unspecified stroke codes) to a main pathological type.(Flynn et al. 2010) The proportion of 'unspecified' stroke cases fell from 67% to 33% when ICD coded data plus natural language processing of scan reports was used, versus ICD coded data alone. Using a physician's classification of radiology reports of 300 randomly selected cases as a reference standard, ICD coding plus analysis of scan reports was more accurate for ischaemic (PPV 95%, 95% CI 90% to 97%) than for haemorrhagic stroke (PPV 77%, 95% CI 69% to 73%).

2.5 Discussion

This was the first systematic assessment of the accuracy of coded hospital, death certificate and primary care data for identifying stroke. Previous published reviews have been less comprehensive in their data inclusion, presentation and analysis, or less precise in their definition of stroke, with the inclusion of TIA, subdural haemorrhage, or all cerebrovascular disease in the reference standard. A previous review based on US studies alone reported similar results but did not include UK based studies or consider either ICD-10 codes or the performance of primary care data or combined data sources.(Andrade et al. 2012) Previous UK-based reviews of ICD or Read code accuracy have reviewed overall accuracy for a wide range of diseases rather than accuracy for stroke specifically,(Thiru, Hassey and Sullivan 2003, Herrett et al. 2010) with limited numbers of stroke/cerebrovascular disease studies.(Khan, Harrison and Rose 2010, Burns et al. 2011, Jordan, Porcheret and Croft 2004, Campbell et al. 2001)

I found wide variation in the performance of ICD codes for stroke and its main types, reflecting the heterogeneity of codes assessed and variation in study settings and methods. These data also show a lack of consensus among stroke epidemiology studies about which codes should be used for identifying stroke outcomes. I have demonstrated that with appropriate selection of stroke-specific codes, PPVs of close to or >90% can be achieved for stroke and each of its main pathological types. Such PPVs will be adequate for many large scale epidemiological studies of the determinants of stroke. However, I found very few studies of the accuracy for stroke of Read coded primary care data or of two or more overlapping data sources. Furthermore, the few available studies of ICD-coded data sources for identification of ischaemic stroke subtypes found that the majority of ischaemic subtype codes were ‘unspecified’,(Goldstein 1998) and reliability of ischaemic subtype classification was limited.(Kessler et al. 1991, Dixon et al. 1998) I found no studies of the accuracy of coded data for identification of subtypes of ICH or SAH.

Within- and between-study comparisons revealed several consistent patterns. First, for stroke of any pathological type, PPV was increased by use of stroke-specific rather than general cerebrovascular codes, making it preferable to use stroke-specific

codes to maximise PPV if no further adjudication of outcomes is planned after identification using ICD codes. Limited evidence suggests that sensitivity is poor when only death certificate data are used as a data source and is markedly increased by including data from hospital admissions, without compromising PPV.(Appelros et al. 2003, Koster et al. 2013a) Results from one study suggest that using general cerebrovascular rather than stroke-specific codes is also likely to increase sensitivity, albeit perhaps by only a small amount and at the expense of a lower PPV.(Ellekjær et al. 1999) To reduce the number of false positives, this method of identifying stroke outcomes is, therefore, probably best used in combination with further steps to confirm which cases are true positives. The best approach for this confirmation process requires further investigation, but could potentially use combinations of ICD codes with coded data from primary care or other sources, or more detailed medical record review. Second, for ischaemic stroke, a greater number of outcomes are identified with little reduction in PPV by using a combination of ischaemic and unspecified stroke codes to identify outcomes. Third, specific codes for ICH and SAH were found to have generally high PPVs (range 71 to 96%). Fourth, across a range of codes for cerebrovascular disease, stroke and pathological stroke types, identification of stroke outcomes using only codes in the primary position increased PPV, but generally by only a modest amount and at the expense of missing true positive outcomes. Furthermore, the relevant studies were of ICD-9 codes only, which are now rarely used outside the USA.(Leibson et al. 1994, Reker et al. 2001, Roumie et al. 2008, Liu et al. 1999, Tirschwell and Longstreth 2002, Benesch et al. 1997, Leone et al. 2004, Rinaldi et al. 2003) Thus, use of appropriately selected codes in both the primary and secondary positions would seem appropriate for most purposes.

There were some limitations. First, since I only searched two online databases, I may have missed a few relevant articles. However, I also reviewed bibliographies of all included publications to increase the sensitivity of our search strategy. Second, my finding that use of the primary diagnostic position improved PPV in some studies may have been due to publication or reporting bias, since many studies did not report on this. Third, since PPV increases with increasing prevalence of the outcome studied, the lower prevalence of ICH and SAH (which together comprise around

20% of all strokes) compared with ischaemic stroke means that the PPVs of these different pathological types are not directly comparable.

Fourth, some included studies had potentially less accurate sources available as a reference standard, such as hospital discharge summaries (a free text summary of the hospital admission, which is often written by less experienced doctors), or non specialist primary care records (potentially based on hospital discharge summaries). I may have overestimated PPV of codes for haemorrhagic stroke types by using such reference standard data from two UK based studies.(Wright et al. 2012, Kirkman et al. 2009) Otherwise, all included studies used potentially more accurate reference standard data sources (independent medical record review and/or expert-led stroke registers). Although I only included studies which used WHO or equivalent definitions of stroke and its main types,(Hatano 1976) there is no 'gold standard' diagnosis for stroke. Indeed, even experts are inconsistent in their ability to diagnose stroke,(Ferro et al. 1996a, Ferro et al. 1998), while choice and timing of imaging (which may vary between centres and therefore between studies) influences the diagnostic accuracy of stroke types.(Wardlaw and Mielke 2005, Fiebach et al. 2002)

Fifth, the paucity of specific published data about the accuracy of Read coded primary care data for stroke is an important further limitation, since up to half of stroke patients are not admitted to hospital in the UK,(Bamford et al. 1986, Rothwell et al. 2004) and hospitalised and non hospitalised strokes may differ in the distribution of pathological types and subtypes and in their risk factor associations.(Schulz and Rothwell 2003) Combining primary care data with other sources (hospital and death certificate data) should improve the detection of non hospitalised cases, reducing potential bias in the selection of cases. Although I identified six systematic reviews of Read code accuracy for a wide range of diseases,(Khan et al. 2010, Thiru et al. 2003, Herrett et al. 2010, Burns et al. 2011, Jordan et al. 2004, Campbell et al. 2001) none included data specifically for stroke. Two excluded studies validated Read codes for cerebrovascular disease,(Mant et al. 2003, Van Staa and Abenhaim 1994) against a reference standard diagnosis of 'cerebrovascular disease'. These 'reference standards' were potentially less accurate because they included hospital ICD codes and patient-self-report without medical

record review, or used internal validation by GP questionnaire (not an independent data source). In addition to improving case ascertainment, primary care data may enhance the sub-classification of potential stroke cases. Around 40% of ICD codes for stroke are of unspecified type, although this proportion may be declining.(http://www.isdscotland.org/ , http://www.saildatabank.com/) Diagnostic codes combined with investigation, procedure, and/or medication codes (in primary care or hospital data) may increase PPV for ischaemic or haemorrhagic stroke.(Gaist et al. 2013, Haesebaert et al. 2013)

2.6 Conclusions

Informed by this review, I recommend using 430, 431, 434, 436 (ICD-9), or I60, I61, I63, I64 (ICD-10), in either the primary or secondary diagnostic position to identify stroke cases with sufficiently high PPV for use in epidemiological studies where further confirmation steps are not envisaged. This may achieve PPVs of >90% for stroke. To increase the number of potential events identified, I suggest using all cerebrovascular disease ICD codes (ICD-9 430-438, or ICD-10 I60-I69, G45, G46) in both primary and secondary positions, but these would have to be combined with additional methods of stroke confirmation to maintain a high PPV. For ischaemic stroke we recommend codes 434, 436 (ICD-9), 433.x1, 434.x1, 436 (ICD-9-CM), and I63, I64 (ICD-10). For haemorrhagic stroke we recommend 430 (ICD-9) and I60 (ICD-10) for SAH, and 431 (ICD-9) and I61 (ICD-10) for ICH. Identifying more detailed stroke subtypes is likely to require coded data from investigations, procedures, and/or drug prescriptions, as well as diagnostic codes, and possibly more detailed review of medical record and imaging data.

Ultimately, UK Biobank aims to improve the accuracy and completeness of stroke outcomes ascertainment by linking multiple sources of coded data. Further work is needed to examine the use of multiple coded data sources to maximise PPV and sensitivity for stroke.

2.7 Appendix: study protocol

2.7.1 Review Question(s)

Primary Question

Accuracy (positive predictive value) of coded health record data (International Classification of Diseases codes, ICD, or Read codes) for stroke and its main pathological types (ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage) using WHO or equivalent definitions in an adult population.

Secondary Questions

A. Sensitivity, specificity, and NPV (negative predictive value) of coded health record data for stroke and its main pathological types (amongst studies where the reference standard is population based).

B. Influence of the range of codes selected (amongst ICD and/or Read codes for 'cerebrovascular disease') on PPV and/or sensitivity for stroke and its main types.

C. Influence of diagnostic position (primary versus secondary) on the PPV and/or sensitivity of ICD codes for stroke and its main pathological types.

D. Accuracy (PPV, sensitivity, specificity, NPV) of multiple sources of coded data (ICD codes from death certificates, and/or ICD codes from hospital records, and/or Read codes from primary care) for stroke and its main pathological types.

2.7.2 Searches

We will search the following databases from 1990 to the date of search:

MEDLINE (Ovid SP); EMBASE (Ovid SP); Clinical Research Practice Datalink; The Health Improvement Network

We will review bibliographies of included publications for any additional relevant articles.

One author will review all titles and abstracts (IG). A second author (RW) will review a 10% sample of titles and abstracts. Two authors will independently review

all potentially relevant full texts and select studies for inclusion (IG and RW). Any areas of uncertainty will be discussed and resolved with a third (CS).

MEDLINE search strategy

1. (cerebrovascular disorders/ or exp basal ganglia cerebrovascular disease/ or exp brain ischemia/ or exp carotid artery diseases/ or exp cerebral small vessel diseases/ or exp intracranial arterial diseases/ or exp intracranial embolism/ and thrombosis/) or exp intracranial hemorrhages/ or stroke/ or exp brain infarction/ or stroke, lacunar/ or vertebral artery dissection/

2. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or appoplex$ or isch?emi$ attack$ or tia$ or SAH).tw.

3. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemisphere$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

4. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraventricular or infratentorial or supratentorial or basal gangli$ or subarachnoid or putaminal or putamen or posterior fossa) adj5 (haemorrhage$ or hemorrhage$ or haematoma$ or hematoma$ or bleed$)).tw.

5. 1 or 2 or 3 or 4

6. medical record/ or medical record review/ or Medical Records Systems, Computerized/ or medical information system/

7. (hospital or GP or medical or general practitioner or health).tw.

8. international classification of diseases/ or Disease/cl [Classification]

9. Clinical Coding/

10. read coding.tw.

11. (ICD-10 or ICD-9 or ICD-9-CM or ICD-10-CM).tw.

12. 6 or 7 or 8 or 9 or 10 or 11

13. exp Sensitivity/ and Specificity/

14. exp Validation Studies/ or exp predictive value/ or exp Reproducibility of Results/

15. (sensitivity or specificity or positive predictive value).tw.

16. (validity or reproducibility).tw.

17. 13 or 14 or 15 or 16

18. 12 and 17

19. 5 and 18

EMBASE search strategy

1. cerebrovascular disease/ or basal ganglion hemorrhage/ or exp brain hematoma/ or exp brain hemorrhage/ or exp brain infarction/ or exp brain ischemia/ or exp carotid artery disease/ or cerebal artery disease/ or cerebrovascular accident/ or exp occlusive cerebrovascular disease/ or exp stroke/

2. stroke unit/ or stroke patient.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

3. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or apoplexy$ or isch?emi$ attack$ or tia$ or SAH).tw.

4. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemisphere$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

5. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraventricular or infratentorial or supratentorial or basal gangli$ or subarachnoid or putaminal or putamen or posterior fossa) adj5 (haemorrhage$ or hemorrhage$ or haematoma$ or hematoma$ or bleed$)).tw.

6. 1 or 2 or 3 or 4 or 5

7. medical record/ or medical record review/ or electronic medical record/ or medical information system/

8. (hospital or GP or medical or general practitioner or health).tw.

9. "international classification of diseases"/ or disease classification/

10. coding/ or "read coding"/

11. read coding.tw.

12. (ICD-10 or ICD-9 or ICD-9-CM or ICD-10-CM).tw.

13. 7 or 8 or 9 or 10 or 11 or 12

14. "sensitivity and specificity"/

15. exp validation study/ or exp predictive value/ or exp reproducibility/

16. (sensitivity or specificity or positive predictive value).tw.

17. (validity or reproducibility).tw.

18. 14 or 15 or 16 or 17

19. 13 and 18

20. 5 and 19

2.7.3 Types of study to be included

We will include studies in adult populations (cohort studies, case control studies, or clinical trials) which compare ICD codes or Read codes for ‘cerebrovascular disease’ against a second source of data for stroke using WHO, or equivalent definitions.

We will exclude studies in highly selected populations with an increased risk of stroke (due to the influence of prevalence on PPV).

Studies are required to report the group(s) of codes validated and the Positive Predictive Value (or data from which this can be calculated).

Studies should use a reference standard of ‘stroke’ (distinguished from transient ischaemic attack or generalised cerebrovascular disease) when calculating PPV, sensitivity, specificity, and NPV.

We will exclude studies which assess < 50 coded events (due to limited precision).

2.7.4 Condition or domain being studied

International Classification of Diseases codes (version 9 or any later version) for cerebrovascular diseases based on hospital admission or death certificate diagnoses.

Read codes (any version) for cerebrovascular diseases.

Studies should publish the group of codes validated (so that the influence of individual code selection on accuracy can be explored).

We will exclude studies of ICD version 8, or earlier primary care systems (e.g. OXMIS codes) because these have been superseded by newer coding systems in the UK.

2.7.5 Participants/population

Any adult population.

2.7.6 Interventions, exposures

Reference standard: We accept that the inter-observer reliability of stroke diagnosis is imperfect, even amongst experts. In the absence of a true 'gold standard' for stroke, we will include studies which use any of the following reference standards for stroke: clinical examination; physician questionnaire; medical record review (primary care and/or hospital records); stroke registers (informed by multiple overlapping data sources, 'hot pursuit', and expert medical record review).

Studies should use a clinical syndrome based definition (WHO, or equivalent) for diagnosing stroke.

2.7.7 Context

The principle criterion is accuracy of coded data (ICD or Read codes for cerebrovascular diseases) for the ascertainment (sensitivity) and confirmation (PPV) of stroke cases in relatively unselected adult populations.

We have pre specified questions to see if PPV and/or sensitivity for stroke and/or its main types are influenced by the groups of cerebrovascular diseases codes selected, or the diagnostic position of these codes (primary versus primary or secondary).

We will exclude studies which include subdural haemorrhage, unspecified cerebrovascular disease and/or TIA in their reference standard definition of stroke.

2.7.8 Outcomes

Primary outcomes

We will calculate Positive predictive value (PPV) of coded data for stroke and/or its main types (ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage) in all included studies using the available published data.

Secondary outcomes

In studies which use population based reference standards, we will calculate sensitivity, specificity, PPV, NPV and stroke prevalence using 2x2 contingency tables.

The reference standard will be grouped into hospital based versus population based according to the following definitions:

Population based: primary care medical records and/or general practitioner questionnaires and/or population based stroke registers used to capture strokes diagnosed out of hospital.

Hospital based: hospital medical records only +/- hospital physician questionnaires which largely capture hospitalised strokes.

If an individual study uses more than one group of cerebrovascular disease codes, more than one diagnostic position (primary versus primary or secondary), or more than one data source (hospital versus death certificate codes versus both), we will explore the influence of code selection, diagnostic position, and code source on PPV/sensitivity for stroke and its main types.

Calculations will be performed independently by two authors (IG and RW) using available published data. Disagreements will be resolved through discussion with a third (CS).

2.7.9 Data extraction (selection and coding)

We will extract data onto study specific proforma.

Covariates of interest:

- Study author

- Publication date
- Country
- Target population age (range)
- Range of ICD/Read codes validated
- Code version (eg. ICD-9, ICD-9-CM, ICD-10)
- Code source (hospital, death certificate or both)
- Number of coded events assessed
- Diagnostic position of codes (primary, secondary or both)
- Diagnosis sought (eg. stroke, ischaemic stroke, haemorrhagic stroke, intracerebral haemorrhage, subarachnoid haemorrhage, or any combination).
- Reference standard used (eg expert hospital record review, population based stroke register)
- Number of coded events confirmed (depending on the diagnosis sought, eg. stroke or one of its main types)
- PPV of selected codes for stroke and/or its main types
- Sensitivity of selected codes for stroke and/or its main types (when the reference standard is population based).

2.7.10 Study-level quality assessment

We will assess methodological quality using a modification of the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2). QUADAS-2 uses fourteen questions to assess study-level risk of bias and generalisability to the target population.

QUADAS-2 was recently modified for a systematic review of the validity of Myocardial Infarction Diagnoses in Administrative Databases.(McCormick et al. 2014) This new version had fewer questions for the assessment of bias and additional questions for the assessment of reporting quality. We will use these questions (because they were developed specifically for studies of the accuracy of coded data), removing the question ‘were the index test results interpreted without knowledge of the results of the reference standard’ because we feel that the codes (index test) are

unlikely to be misinterpreted, and adding the question 'was the study UK-based (?)' to assess generalisability to the UK population.

Our modified QUADAS-2 will therefore include five questions for the assessment of reporting quality, three for assessment of generalisability, and six for assessment of risk of bias (fourteen in total). Each question will score 'low', 'high', or 'unclear' reporting quality/generalisability/risk of bias according to the rules (below).

An overall quality score (0-14) will be derived for each study by adding the number of questions which scored 'high reporting quality', 'high generalisability', and 'low risk of bias'.

2.7.11 Modified QUADAS-2 scoring methods

1) Assessment of reporting quality

a) Were the selection criteria clearly described?

High reporting quality: population selection criteria clearly described.

Low reporting quality: population selection criteria unclear.

b) Was execution of the index test described in sufficient detail to allow replication of the test? (The index test was the coding algorithm used in the study.)

Studies were excluded from this review if the codes used were not reported.

High reporting quality: diagnostic position of codes (primary vs. primary or secondary diagnoses) clearly reported.

Low reporting quality: diagnostic position of codes not clearly reported.

c) Was execution of the reference standard described in sufficient detail to permit its replication?

High reporting quality: reported the information used to make the reference standard diagnosis (eg. full medical record vs. abstracted data, brain imaging to confirm stroke subtypes) and described the adjudicators' expertise (eg. expert stroke physician).

Low reporting quality: did not report the information used to make the reference standard diagnosis and/or did not report adjudicators' expertise.

d) If participants were excluded from the final analysis, were they described and were the reasons for this exclusion explained?

High reporting quality: reported the numbers excluded from the final analyses and the reason(s), or there were no exclusions.

Low reporting quality: did not report the numbers excluded from the final analyses and the reason(s).

e) Were uninterpretable/intermediate results reported?

High reporting quality: 'uncertain' diagnoses were reported (eg. if there was 'insufficient information in the medical record to make a diagnosis, and/or it was reported if 'possible strokes' were included in the 'confirmed stroke' category.)

Low reporting quality: no reporting of 'uncertain' results.

2) Assessment of generalisability (to UK population)

a) Country: was the study UK based?

High generalisability: UK based population.

Low generalisability: non UK based population.

b) Population selected: was the spectrum of patients selected representative of the patients who will receive the diagnosis in practice?

High generalisability: the study included patients diagnosed and treated in a representative mixture of specialist and non specialist settings, and the population was otherwise relatively unselected.

Low generalisability: the study was performed in a more selected population (eg. restricted to patients admitted to a specialist stroke unit, where coding performance might be higher).

Unclear: insufficient published information.

c) Reference standard: were the same clinical data available when test results were interpreted as would be available when used in practice?

High generalisability: medical record data (extracts or full record) including brain imaging data (original scans/written reports). Brain imaging would be used in current practice to exclude stroke mimics/diagnose stroke subtypes.

Low generalisability: brain imaging (original scans/reports) not available.

Unclear generalisability: insufficient published information.

3) Assessment of risk of bias

a) ***Selection bias: did the whole sample, or a random selection of the sample receive verification by the reference standard?***

Low risk of bias: The whole sample/random selection of the sample received verification using medical records/physician questionnaire.*

High risk of bias: some of the sample did not receive verification because reference standard data were missing* (missing records/unreturned questionnaires).

Unclear risk of bias: insufficient information published.

*It is assumed that if the reference standard was a prospectively generated, expert led, population based stroke register, which used multiple sources of case ascertainment and confirmation, that this reference standard is 'complete' (ie. it is unlikely to have missed 'true positive' stroke cases).

b) Blinding: were the reference standard results interpreted without knowledge of the results of the index test?

Low risk of bias: blinding present, or the reference standard diagnosis was made prior to the study (eg., the reference standard was a prospectively generated stroke register)

High risk of bias: blinding not present, or not reported.

c) Independence: was the reference standard independent of the index test?

Low risk of bias: the reference standard was independent of the index test.

High risk of bias: the index test formed part of the reference standard (eg. coded diagnoses were used to identify stroke cases for a population based register, and there was no further confirmation).

d) Differential verification: did all patients receive the same reference standard regardless of the index test result?

Low risk of bias: yes

High risk of bias: some/all code positive cases received different reference standards from code negative cases eg. stroke code positive cases not present in a stroke register (potential false positive cases) were selected for subsequent medical record review, but stroke code negative cases present in

stroke register (potential false negative cases) did not have subsequent medical record review.

e) Reference standard: is the reference standard likely to correctly classify the target condition?

Low risk of bias: the reference standard was likely to identify all hospital admitted strokes (plus strokes in the community for death certificate codes), AND, was either based on an expert (neurologist or stroke physician) reviewing the full medical record, or was based on a non expert (eg. research assistant, research nurse, or 'adjudicator') following clearly described rules (eg. using CT reports to exclude stroke mimics or using CT reports to differentiate haemorrhage from ischaemic stroke, where applicable).

High risk of bias: the study validated death certificate codes and the reference standard was not population based (and therefore risked misclassifying true strokes diagnosed out of hospital as false positive codes), and/or the diagnosis was made by a non expert and there was not a clear protocol to exclude stroke mimics, or to differentiate haemorrhagic from ischaemic stroke.

Unclear: insufficient published data.

f) Timing: was the time period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?

Low risk of bias: The information used to make the reference standard diagnosis was the same as the information used at the time of coding.

High risk of bias: The information used to make the reference standard diagnosis was not the same as the information used at the time of coding.

Unclear: insufficient published data.

2.7.12 Strategy for data synthesis

We will cross classify coded stroke diagnoses with the reference standard diagnosis ('stroke' versus 'non stroke').

PPV (%) = true positive codes / [true positive codes + false positive codes] x 100

NPV (%) = true negative codes / [true negative codes + false negative codes] x 100

True positive codes = coded diagnosis ‘stroke’ and reference standard ‘stroke’.

False positive codes = coded diagnosis ‘stroke’ and reference standard ‘non stroke’.

True negative codes = coded diagnosis ‘non-stroke’ and reference standard ‘non stroke’.

False negative codes= coded diagnosis ‘non stroke’ and reference standard ‘stroke’.

Where the reference standard is population based, we will construct standard 2x2 tables describing binary test results (coded diagnosis ‘stroke’ and coded diagnosis ‘non stroke’) cross classified with binary reference standard results (‘stroke’ and ‘non stroke’). We will use this data to calculate sensitivity, specificity, PPV, NPV, and 95% confidence intervals. We will tabulate results for visual inspection to assess the influence of individual code selection, diagnostic position, and code source on the PPV, sensitivity, specificity and NPV of coded data for stroke/its main types. Where possible, and to limit the impact of between study heterogeneity, we will use within study as well as between study comparisons.

We will assess heterogeneity between studies by inspection of tabulated data.

We will not quantify publication bias as there is no assessment applicable to test accuracy.

2.7.13 Dissemination plans

We will present our findings at local, national and international meetings. We plan to publish a full paper in a peer reviewed scientific journal.

Chapter 3 Accuracy of patient self-report of stroke: a systematic review

- The UK Biobank baseline assessment included participant self-report of stroke.
- Self-report may be a useful method to ascertain and confirm potential stroke cases in UK Biobank.
- In this chapter, I report the results of a systematic review of the accuracy of patient self-report of stroke.
- I used a comprehensive search strategy, critically appraised study quality, and assessed factors affecting self-report accuracy.
- I found that characteristics of the study population strongly influenced reporting accuracy.
- In populations with low stroke prevalence (<10%), a large proportion of self-reported strokes (~1/3 to ~3/4) were false positive.
- I concluded that self-report was unlikely to be helpful in identifying stroke cases in UK Biobank without further confirmation.

3.1 Introduction

Self-report of stroke may be a useful method to ascertain and confirm potential stroke cases among UK Biobank participants. Participants with confirmed stroke before or at the time of UK Biobank recruitment would have 'prevalent stroke'. Participants who go on to develop stroke during follow-up would either have 'recurrent stroke' (if they also had a past history of stroke), or 'incident stroke' (if their first-in-a-lifetime stroke event occurred during UK Biobank follow-up). For large prospective studies, like UK Biobank, disease outcomes need to be identified with adequate sensitivity (the proportion of true stroke cases which are identified by participant self-report), and with adequate Positive Predictive Value, PPV (the proportion of self-reported cases which are true cases of stroke). PPV depends on sensitivity, specificity (the number of true negative non stroke controls identified),

and stroke prevalence. Table 3.1 shows how sensitivity, specificity, PPV, NPV, and stroke prevalence are calculated.

Table 3.1 Calculation of Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and stroke prevalence*.

	Reference standard 'stroke'	Reference standard 'no stroke'	
Self-report 'stroke'	True positive (TP)	False positive (FP)	PPV = $\frac{TP}{(TP+FP)}$
Self-report 'no stroke'	False negative (FN)	True negative (TN)	NPV = $\frac{TN}{(FN+TN)}$
	Sensitivity= $\frac{TP}{(TP+FN)}$	Specificity = $\frac{TN}{(FP+TN)}$	

*Stroke prevalence is calculated by (TP+FN) / (TP+FP+FN+TN)

Maximising PPV will limit the number of false positive cases included in future nested case cohort or case control studies in UK Biobank of the associations between risk factors, stroke, its main types, and subtypes. Maximising PPV will, in turn, minimise loss of statistical power through misclassification of cases. False negatives can be tolerated more easily in nested case cohort/case control studies because they are diluted by the larger control population (therefore the focus for UK Biobank is on maximising PPV and minimising false positives).

3.1.1 Self-report of stroke in UK Biobank

At recruitment to UK Biobank, participants completed a detailed questionnaire which included the question 'Has a doctor ever told you that you have had a stroke?' Positive responses or unclear answers were confirmed in a brief nurse led interview. Based on this questionnaire, the prevalence of stroke in UK Biobank at recruitment was ~1.4%.(http://biobank.ctsu.ox.ac.uk/crystal) The British Heart Foundation estimated UK stroke prevalence up to 2010, based on data from general practice records. (Townsend et al. 2012) Stroke prevalence was between 2 and 3.5% depending on region (England, Scotland, Wales), and gender. Prevalence of stroke in UK Biobank is expected to be lower than UK wide stroke prevalence due to the fact that study participants are healthier than the population as a whole ('healthy cohort

effect').(Table 1.1, Lindsted et al. 1996) UK Biobank aims to use multiple overlapping sources of data to identify potential stroke cases (maximising sensitivity), and to confirm true cases of stroke, and/or its pathological types (maximising PPV). Potential self-reported cases are likely to need further confirmation in order to maximise PPV. There have been no previous systematic reviews of the accuracy of patient self-report of stroke. It is known that self-report accuracy varies according to the disease reported.(Harlow and Linet 1989)

3.2 Aims

To assess the potential contribution of participant self-report to the identification, and/or confirmation of stroke in UK Biobank (prevalent and/or incident cases), I performed a systematic review of published studies assessing the accuracy of patient self-report of stroke against a reference standard for stroke (using WHO, or equivalent definitions). I focussed on PPV, but also recorded, where possible, the sensitivity, specificity, and NPV of self-report.

3.3 Methods

A detailed study protocol is presented in Appendix 3.7

3.3.1 Search strategy

Using a combination of medical subject heading and text word terms for 'stroke', 'self-report', 'accuracy', 'medical records' and 'diagnosis', I searched Medline and Embase to November 2013 for studies assessing the accuracy of self-report of stroke against a reference standard diagnosis of stroke.(Appendix 3.7.2) I also searched the Cochrane Database of Systematic Reviews for relevant reviews of diagnostic test accuracy of stroke self-report. Bibliographies of included publications were reviewed to identify any additional relevant studies. I assessed eligibility by reviewing all titles and abstracts, and the full text of potentially relevant articles, and resolved any uncertainties through discussion with my supervisor, Professor Cathie Sudlow.

3.3.2 Eligibility criteria

I included studies which assessed the accuracy of patient self-report of stroke (with or without transient ischemic attack [TIA], or synonyms for either) against a reference standard diagnosis of stroke. I hypothesised that asking about TIA (or its synonyms) might increase sensitivity for stroke, and so I also included studies which compared self-report of stroke or TIA against a reference standard of stroke. Included studies had to report the method of self-report, the reference standard used (any combination of hospital/primary care medical record review, hospital/primary care physician questionnaire, expert clinical examination, or hospital/population based stroke registers), and the positive predictive value (PPV) +/- sensitivity, specificity, negative predictive value (NPV) of self-report (or provide data from which these values could be calculated, as shown in Table 3.1, above). I excluded studies which assessed self-report of 'cerebrovascular disease', 'symptoms' or 'past medical history' unless stroke was specifically mentioned. I also excluded studies which used only coded data (e.g. International Classification of Diseases codes) as the reference standard to confirm cases, studies which did not distinguish confirmed

stroke cases from transient ischaemic attack (TIA) or other cerebrovascular disease, and, to improve precision, studies with <50 self-reported strokes.

3.3.3 Data extraction

I extracted information from each included study on: the nature of the population surveyed (country, age range, selection criteria); number of participants included and response rate (proportion of potential participants who agreed to take part and completed questionnaires or attended interviews); question(s) asked (stroke, or stroke plus TIA/synonyms); mean (or median) age at self-report; recall period (years or lifetime); reference standard(s) used and source of data ('hospital' which includes only hospital diagnosed strokes, or 'population' which includes strokes diagnosed in the community); presence or absence of missing data; presence or absence of blinding of adjudicators (physicians or researchers who established the reference standard diagnosis) to participant self-report; presence or absence of differential verification (use of different reference standards for self-report positive versus self-report negative responses); PPV, sensitivity, specificity and NPV of self-report; number of reports of stroke which were confirmed TIAs.

3.3.4 Data analysis

I tabulated results for visual inspection to assess factors which might influence the accuracy of self-report including: estimated stroke prevalence; age at self-report (mean or median); recall time (years); question asked (stroke or stroke plus TIA/synonyms). Where possible, I used within study comparisons to assess the influence of age, recall time and question(s) asked on the accuracy of self-report. The denominator population was the final number of participants (excluding non responders) for whom complete reference standard data were available. Stroke prevalence was the number of 'true strokes' (TP+FN) divided by the denominator population (TP+FP+FN+TN). I only calculated sensitivity, specificity and stroke prevalence when the reference standard was population based (i.e. included general practitioner medical records, general practitioner questionnaires, or physician assessment of all participants, to capture strokes diagnosed out of hospital).
I assessed risk of bias at individual study level using the revised Quality Assessment of Diagnostic Studies tool (QUADAS-2), (Whiting et al. 2011) but did not exclude

studies on the basis of bias assessments. Risk of bias was scored as 'low', 'high', or 'unclear' in response to specific questions which considered patient selection (study design, sampling methods, exclusion criteria); index test (self-report questionnaire design, blinding to the reference standard); reference standard (source of data, blinding to self-report status); flow and timing (participant response rates, missing reference standard data, presence of differential verification). The study protocol (Appendix 3.7.11) provides a detailed list of questions and scoring methods.
I calculated 95% confidence intervals for PPV, sensitivity, specificity, and NPV in *Stata* version 12 using the Wilson method for binomial proportions. I did not undertake formal meta-analysis or meta-regression due to the heterogeneity between studies in their methods, participant characteristics and reporting, and because the number of studies available for inclusion in any potential meta-regression analysis was small (<10).(Higgins 2011)

3.4 Results

From 1707 publications identified, I reviewed 148 full texts, and eventually included 17 studies (Figure 3.1)

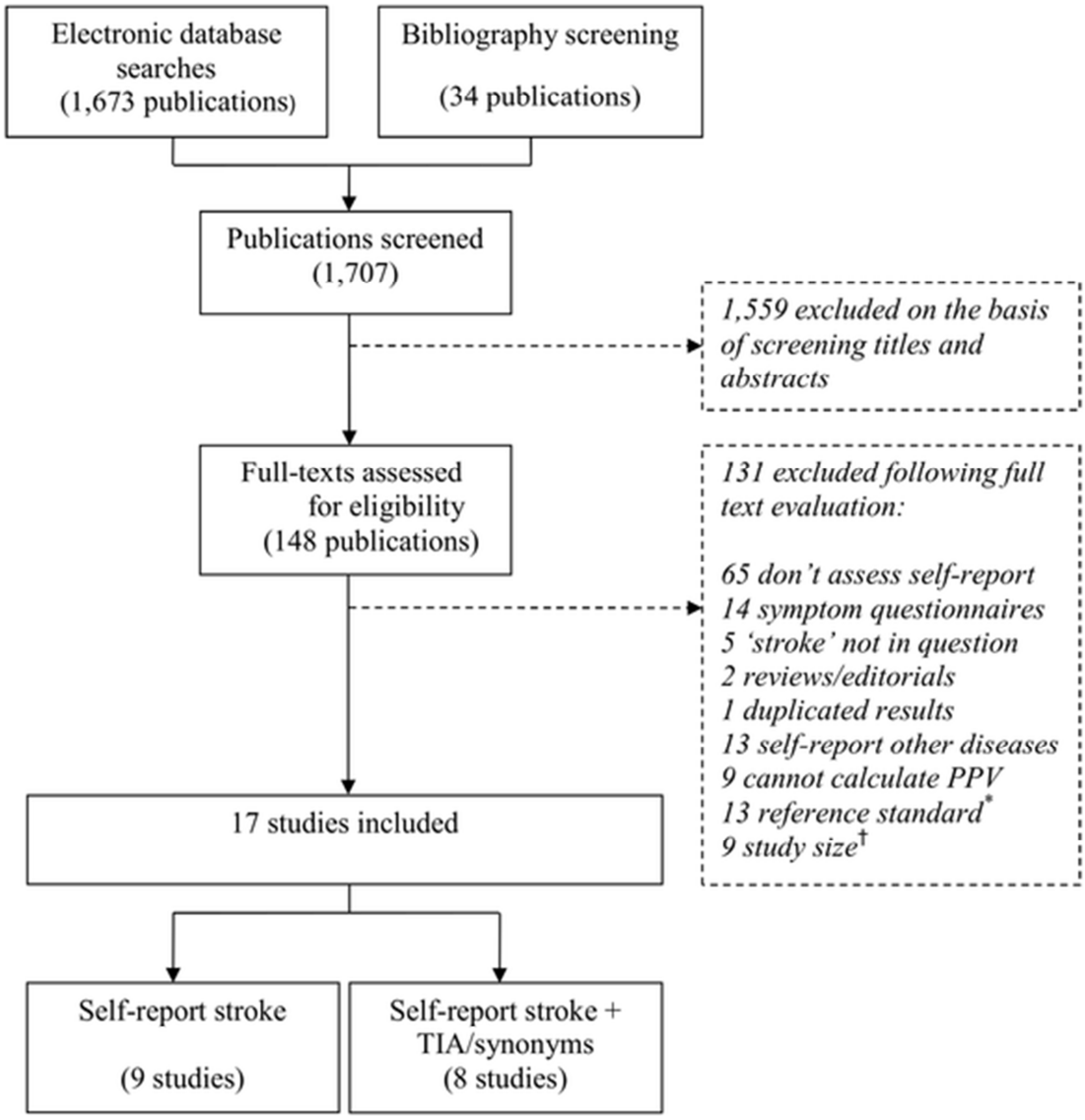

Figure 3.1Study selection flow diagram

*Reference standard is cerebrovascular disease, or includes TIA, or uses ICD codes for stroke.

†<50 self-reports of stroke validated, or number unpublished.

3.4.1 Characteristics of included studies

Characteristics of included studies are displayed in Table 3.2. Studies were from the UK,(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995) elsewhere in Europe, (Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Bots et al. 1996) Japan,(Yamagishi et al. 2009) North America,(Simpson et al. 2004, Jin et al. 2010, Colditz et al. 1986, Okura et al. 2004, Heckbert et al. 2004, Bergmann et al. 1998) Australia,(Barr et al. 2009) and New Zealand.(Teh et al. 2013) Potential participants were selected based on geographic location and/or age in all studies. The age of potential participants ranged from >20 to >80 years old. Additional selection criteria in some studies included gender, (Walker et al. 1998, Simpson et al. 2004, Colditz et al. 1986, Heckbert et al. 2004) occupation,(Britton et al. 2012, Colditz et al. 1986) presence or absence of disability, (Kriegsman et al. 1996, Simpson et al. 2004) absence of moderate/severe cognitive impairment,(Walker et al. 1998, Simpson et al. 2004) place of residence,(Kriegsman et al. 1996) or medication use (non-steroidal anti-inflammatory drug prescription).(Fourrier-Reglat et al. 2010)

Responses were ascertained by post,(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Engstad 2000, Yamagishi et al. 2009, Colditz et al. 1986, Okura et al. 2004, Heckbert et al. 2004) during routine outpatient visits,(Machon et al. 2013) by face-to-face interview,(Kriegsman et al. 1996, Simpson et al. 2004, Jin et al. 2010, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009, Teh et al. 2013) or by telephone.(Machon et al. 2013) Response rates were ≥80% in five studies,(Walker et al. 1998, O'Mahony et al. 1995, Machon et al. 2013, Kriegsman et al. 1996, Barr et al. 2009)] 60–79% in six,(Britton et al. 2012, Bots et al. 1996, Yamagishi et al. 2009, Simpson et al. 2004, Jin et al. 2010, Bergmann et al. 1998) and <60% in three (23–57%).(Fourrier-Reglat et al. 2010, Okura et al. 2004, Teh et al. 2013) The remaining three studies did not report response rates.(Engstad 2000, Colditz et al. 1986, Heckbert et al. 2004) Four studies compared characteristics of responders to non responders.(O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Okura et al. 2004, Barr et al. 2009) The two largest studies (including 10,000 and 120,000 potential participants) found that responders were older than non responders and more often female,(Fourrier-Reglat et al. 2010,

Barr et al. 2009) but absolute differences were small (mean age 2 or 3 years higher, 1 or 6% more females).

Six studies asked participants to report ‘stroke’,(Walker et al. 1998, Bots et al. 1996, Simpson et al. 2004, Jin et al. 2010, Barr et al. 2009, Teh et al. 2013) five to report ‘stroke, mini-stroke or transient ischemic attack (TIA)’,(Britton et al. 2012, O'Mahony et al. 1995, Okura et al. 2004, Heckbert et al. 2004, Bergmann et al. 1998) and three to report stroke, but including stroke synonyms in the question (cerebral haemorrhage/brain haemorrhage/infarction/thrombosis/subarachnoid haemorrhage).(Machon et al. 2013, Engstad 2000, Yamagishi et al. 2009) All but three studies published the specific question(s) asked.(Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Colditz et al. 1986) The period of recall ranged from six months to 5 years in five studies,(Britton et al. 2012, Jin et al. 2010, Colditz et al. 1986, Heckbert et al. 2004, Barr et al. 2009) 10 to 22 years in three studies,(Walker et al. 1998, Yamagishi et al. 2009, Bergmann et al. 1998) and was lifetime in the remaining nine.(O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Bots et al. 1996, Simpson et al. 2004, Okura et al. 2004)

Table 3.2:Characteristics of included studies

	Patient selection*			Index test† (self-report)			Reference standard		Flow and timing		
Study	**Country**	**Target population age (range)**	**Population included**	**Self-report method**	**Question(s) asked**	**Recall period‡**	**Source of data§**	**Blind to self-report¶**	**Response rate** (%)**	**Missing data††**	**Differential verification**
Reglat	France	> 20	Prescribed NSAIDs‡‡	Postal survey	-	Lifetime	Population§§	Yes	23¶¶	Yes	No
Yamagishi	Japan	40-69	-	Postal survey	Stroke+3	10	Population***	No***	77	No	Yes***
Walker	UK	40-59	Men†††	Postal survey	Stroke	12-14	Population‡‡‡	No‡‡‡	90	Yes	Yes‡‡‡
Kriegsman	Netherlands	55-85	Independent§§§	Interview	-	Lifetime	Population§§	Yes	82	Yes	No
Simpson	US	> 65	Disabled women¶¶¶	Interview	Stroke	Lifetime	Population	Unclear	71	Yes	Unclear
Jin	Canada	> 65	Cognitive impairment****	Interview	Stroke	1	Population	No	64	Yes	Unclear
Engstad	Norway	> 24	-	Postal survey	Stroke+1	Lifetime	Population	No††††	Unclear	Yes	No
Barr	Australia	>25	-	Interview	Stroke	5	Hospital	Yes	82	Yes	No
Bots	Netherlands	> 55	-	-	Stroke	Lifetime	Population	Unclear	78	Yes	Unclear
Britton	UK	35 -55	Civil servants	Postal survey	Stroke/TIA	3 - 4	Population‡‡‡‡	No‡‡‡‡	66 - 70	No	Yes‡‡‡‡
Colditz	US	30-55	Female nurses	Postal survey	-	2	Hospital	Yes	Unclear	Yes	No

	Patient selection*			Index test† (self-report)			Reference standard		Flow and timing		
Teh	New Zealand	>80	-	Interview	Stroke	Lifetime	Population	Yes	56	Yes	No
Machon	Spain	29-70	-	Questionnaire/ Interview	Stroke+2	Lifetime/3§§§§	Population	Unclear	99	No	No
O'Mahony	UK	>45	-	Postal survey	Stroke+5/TIA	Lifetime	Population	No¶¶¶¶	83	Yes	Yes¶¶¶¶
Heckbert	US	50-79	Postmenopausal women*****	Interview/ Postal survey	Stroke/TIA	0.5 - 1	Hospital	Unclear	Unclear	Unclear	No
Okura	US	>45	-	Postal survey	Stroke/TIA	Lifetime	Population	Unclear	47	No	No
Bergmann	US	25-74	-	Interview	Stroke/TIA	12-22	Hospital	Unclear	78	Yes	No

Stroke/TIA: Has a doctor ever told you that you had a stroke, mini stroke or Transient Ischemic Attack (TIA)?

Stroke +1: Do you have or have you ever had a stroke or cerebral haemorrhage?

Stroke +2: Have you ever been told by a physician that you have or have had a stroke, cerebral thrombosis, or cerebral haemorrhage?

Stroke+3: Have you ever been told by a physician that you had a stroke, cerebral haemorrhage, cerebral infarction, or subarachnoid haemorrhage?

Stroke+5: Have you ever had a stroke, cerebral haemorrhage, cerebral thrombosis, brain haemorrhage, subarachnoid haemorrhage, or cerebrovascular accident?

*The majority of included studies randomly sampled participants from within their chosen sampling frame.

†None of the studies reported if self-report was interpreted blind to the reference standard diagnosis.

‡Recall period for self-report of incident stroke. 'Lifetime'= self-report of stroke at any time prior to recruitment to a study.

§Source of reference standard data. Population: primary care medical records +/- general practitioner questionnaires +/- population based stroke registers +/- clinical examination of all participants. Hospital: hospital based records only +/- hospital physician questionnaire.

[¶]Yes: the reference standard diagnosis was made blind to self-report status. Unclear: not clearly reported. No: the reference standard diagnosis was made unblind to self-report status, or blinding was jeopardised (history and examination of participants, or differential verification).

[**] Proportion of potential participants who agreed to take part, completed and returned questionnaires, or attended interviews.

[††] Yes: participants excluded from the final analysis due to missing reference standard data, or because the reference standard diagnosis was 'unclear'. Unclear: insufficient information published. No: reference standard data complete for all participants.

[‡‡]Prescribed Celecoxib, Rofecoxib, or traditional non-selective Non-Steroidal Anti-Inflammatory Drugs (NSAIDs).

[§§]General Practitioner questionnaire only.

[¶¶]Amongst these participants ~58% of GPs responded with questionnaire data.

[***]Population stroke register. Medical records were re-examined if a patient reported stroke but the register was negative.

[†††]Excluded those with 'severe mental or physical disability'.

[‡‡‡]General Practitioner questionnaire. Primary care records were re-examined for evidence of stroke if a patient reported stroke but the GP did not (apparent false positive cases).

[§§§]Excluded residential or nursing home residents.

[¶¶¶]Excluded participants with moderate or severe cognitive impairment.

[****] Final sample included participants with cognitive impairment and a random sample without cognitive impairment.

[††††]In this study, physicians took a history from patients. Results of history and physical examination were used, in addition to medical record review, to determine the final diagnosis (presence or absence of stroke).

[‡‡‡‡] Source of the reference standard was coded hospital data plus medical record review or GP questionnaire. If self-report or coded data were positive for stroke, a committee reviewed data abstracted from medical records. If self-report was positive for stroke and hospital data were unavailable, GP questionnaire was requested.

[§§§§]Self-report included prevalent events (baseline questionnaire) plus additional prevalent or incident events (3 year follow-up telephone interview).

[¶¶¶¶]Validation included physician history and examination of patients who self-reported stroke. The final reference standard diagnosis was made by expert consensus using multiple sources of information. Good medical records took precedence over the physician diagnosis at home visit.

[*****]Includes clinical trial participants on different HRT regimens, low fat diet, and calcium and vitamin D supplementation

The reference standard was population based (primary care data, primary care plus hospital data, and/or clinical examination of all participants) in thirteen studies,(Okura et al. 2004, Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Bots et al. 1996) and hospital based (hospital data only) in four.(Colditz et al. 1986, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009) Stroke was confirmed by general practitioner questionnaire,(Fourrier-Reglat et al. 2010, Kriegsman et al. 1996) medical record review,(Machon et al. 2013, Bots et al. 1996, Colditz et al. 1986, Okura et al. 2004, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009) presence on a stroke register plus medical record review,(Yamagishi et al. 2009) clinical examination plus medical record review,(Engstad 2000) or a combination of these methods.(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Simpson et al. 2004, Jin et al. 2010, Teh et al. 2013) Stroke prevalence ranged from 0.1%-17% in ten studies which used a population based reference standard and published sufficient data to estimate prevalence.(Walker et al. 1998, O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Yamagishi et al. 2009, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013)

3.4.2 Assessment of bias

Detailed results of the bias assessment are displayed in Table 3.3. All studies had 'high' or 'unclear' risk of bias in at least one category. Incomplete reference standard data (missing or irretrievable records),(Walker et al. 1998, O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Bots et al. 1996, Simpson et al. 2004, Jin et al. 2010, Colditz et al. 1986, Bergmann et al. 1998, Barr et al. 2009, Teh et al. 2013) absence of blinding of adjudicators to self-report status, (Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Engstad 2000, Yamagishi et al. 2009, Jin et al. 2010), and participant response rates (<80%),(Britton et al. 2012, Fourrier-Reglat et al. 2010, Bots et al. 1996, Yamagishi et al. 2009, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013, Bergmann et al. 1998) were the most frequent reasons for 'high risk' of bias.

Only five studies reported that adjudicators were blind to participant self-report results.(Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Colditz et al. 1986, Barr et al. 2009, Teh et al. 2013) In six, presence or absence of blinding was not clearly reported.(Machon et al. 2013, Bots et al. 1996, Simpson et al. 2004, Okura et al. 2004, Heckbert et al. 2004, Bergmann et al. 1998) In one study, the reference standard diagnosis was made following physician examination of patients, unblinded to self-report status.(Jin et al. 2010) In five other studies blinding was jeopardised because the reference standard included history and examination of patients,(O'Mahony et al. 1995, Engstad 2000) or because records of apparent false positive reports were re-examined for evidence of stroke.(Walker et al. 1998, Britton et al. 2012, Yamagishi et al. 2009) In one of these studies, re-examination of records of apparent false-positive reports led to confirmation of a few additional stroke cases, and increased the PPV of patient self-report from 41% (95% CI 35–48) to 56% (95% CI 49–62).

The self-report method most often scored ‘unclear’ risk of bias. Three studies did not publish the specific question(s) asked,(Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Colditz et al. 1986) and eight (which used face-to-face interviews) did not report presence or absence of blinding of the interviewer to the reference standard diagnosis.(Machon et al. 2013, Simpson et al. 2004, Jin et al. 2010, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009, Teh et al. 2013)

Table 3.3 Study-level risk of bias using the Quality Assessment of Diagnostic Studies tool (QUADAS-2).

STUDY	PATIENT SELECTION			INDEX TEST		REFERENCE STANDARD		FLOW AND TIMING		
	Sampling method*	Study design†	Population included‡	Interpreted blind§	Threshold specified¶	Source**	Blind to self-report††	Response rate‡‡	Missing data§§	Differential verification¶¶
Reglat	☺	☺	☺	?	?	☺	☺	☹	☹	☺
Yamagishi	☺	☺	☺	☺	☺	☺	☹	☹	☺	☹
Walker	☺	☺	☹	☺	☺	☺	☹	☺	☹	☹
Kriegsman	?	☺	☹	☹	?	☺	☺	☺	☹	☺
Simpson	☺	☺	☹	?	☺	☺	?	☹	☹	?
Jin	?	☺	☹	?	☺	☺	☹	☹	☹	?
Engstad	☺	☺	☺	☺	☺	☺	☹	?	☹	☺
Barr	☺	☺	☺	?	☺	☹	☺	☺	☹	☺
Bots	☺	☺	☺	?	☺	☺	?	☹	☹	?
Britton	☺	☺	☺	☺	☺	☺	☹	☹	☺	☹
Colditz	?	☺	☺	?	?	☹	☺	?	☹	☺
Teh	?	☺	☺	?	☺	☺	☺	☹	☹	☺
Machon	☺	☺	☺	?	☺	☺	?	☺	☺	☺
O'Mahony	☺	☺	☺	☺	☺	☺	☹	☺	☹	☹
Heckbert	☺	☺	☺	?	☺	☹	?	?	?	☺
Okura	☺	☺	☺	☺	☺	☺	?	☹	☺	☺
Bergman	☺	☺	☺	?	☺	☹	?	☹	☹	☺

☺ Low risk of bias
☹ High risk of bias
? Unclear bias risk

Rules for assessment of bias are displayed in Appendix.3.7.11.

[*]Was consecutive or random sampling used?

[†]Was a case-control design avoided?

[‡]Were inappropriate exclusions avoided?

[§]Was self-report interpreted blind to the reference standard diagnosis?

[¶]Was the test threshold (self-report positive versus self-report negative) pre specified?

[**]Is the reference standard likely to correctly classify the target condition?

[††]Was the reference standard diagnosis made blind to participant self-report status?

[‡‡]Were all participants included in the analysis?

[§§]Did all participants receive a reference standard?

[¶¶]Did all participants receive the same reference standard?

Other sources of bias included use of hospital versus population based reference standards,(Colditz et al. 1986, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009) exclusion of particular types of participants (e.g. based on cognitive impairment, severe disability, or residence in a nursing home),(Walker et al. 1998, Kriegsman et al. 1996, Simpson et al. 2004, Jin et al. 2010) and differential verification of the reference standard (different data used to verify self-report positive versus negative cases).(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Yamagishi et al. 2009) Most studies used a source of data (primary care records, general practitioner questionnaire, examination of all participants) which captured strokes diagnosed out of hospital.(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Yamagishi et al. 2009, Engstad 2000, Bots et al. 1996, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013) The remaining studies, which used hospital-based reference standards,(Colditz et al. 1986, Heckbert et al. 2004, Bergmann et al. 1998, Barr et al. 2009) had a higher risk of bias due to the potential for missing 'true' stroke cases diagnosed out of hospital.

3.4.3 Accuracy of self-report

PPV of self-report ranged from 22–87%. Among ten studies which used a population based reference standard, and had sufficient published data, sensitivity of self-report varied (from 36–98%), but specificity and NPV were consistently high (from 88 to 99.9%) (Table 3.4).(O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Yamagishi et al. 2009, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013) The proportion of self-reported strokes which were not strokes but confirmed to be TIAs (according to the reference standard) ranged from 6–25% among six studies with available relevant data.(Walker et al. 1998, O'Mahony et al. 1995, Machon et al. 2013, Bots et al. 1996, Colditz et al. 1986, Teh et al. 2013) In these studies, if these confirmed TIAs were considered to be true rather than false positive stroke cases, the revised PPV was >75% in all but one study (Table 3.5).

3.4.4 Factors influencing accuracy

The range of PPVs for self-report was similar in the studies with very low response rates (<60%),(Fourrier-Reglat et al. 2010, Okura et al. 2004, Teh et al. 2013) to that of studies with higher response rates (≥60%).(Walker et al. 1998, O'Mahony et al. 1995, Machon et al. 2013, Kriegsman et al. 1996, Yamagishi et al. 2009, Engstad 2000, Bergmann et al. 1998) Overall, the range of PPVs appeared similar in blinded, (Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Colditz et al. 1986, Barr et al. 2009) versus unblinded studies,(Walker et al. 1998, Jin et al. 2010) and in population based,(O'Mahony et al. 1995, Machon et al. 2013, Fourrier-Reglat et al. 2010, Kriegsman et al. 1996, Engstad 2000, Yamagishi et al. 2009, Bots et al. 1996, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013) versus hospital based studies.(Colditz et al. 1986, Heckbert et al. 2004, Bergmann et al. 1998, Bots et al. 1996)

Table 3.4 PPV, sensitivity, specificity and NPV of self-report[*] (all included studies).

Study	Study population[†] TP+FP+FN+TN	Self-report stroke TP+FP	Self-report confirmed TP	Total stroke TP+FN	Stroke prevalence (%) (TP+FN)/(TP+FP+FN+TN)	PPV (% & 95% CI) TP/(TP+FP)	Sensitivity (% & 95% CI) TP/(TP+FN)	Specificity (% & 95% CI) TN/(FP+TN)	NPV (% & 95%CI) TN/(TN+FN)
Reglat	16935	248	156	227	1.3	63 (57-69)	69 (62-74)	99 (99.3-99.6)	99.6 (99.5-99.7)
Yamagishi	89914	1848	1051	1447	1.6	57 (55-59)	73 (70-75)	99.1 (99.0-99.2)	99.5 (99.5-99.6)
Walker	5907	201	112	126	1.9	56 (49-62)	89 (82-93)	97 (96.8-97.7)	99.8 (99.6-99.9)
Kriegsman	2380	119[‡]	69[‡]	119[‡]	5.0	58 (49-66)	58 (49-66)	98[‡] (97.1-98.3)	97.8 (97.1-98.3)
Simpson	945	94	67	68	8.0	71 (61-79)	98 (92-99)	97 (95.7-97.9)	96.9 (95.5-97.9)
Jin	1536	113	92	184	11.9	81 (73-88)	50 (43-57)	98 (97.6-98.9)	93.5 (92.1-94.7)
Engstad[§]	17122[§]	269	213	-[§]	-[§]	79 (74-84)	-[§]	-[§]	-[§]
Barr	-	87	33	-	-	38 (29-48)	-	-	-
Bots[¶]	-[¶]	285	191	-[¶]	-[¶]	67 (61-72)	-[¶]	-[¶]	-[¶]
Britton[¶]	-[¶]	106	83	-[¶]	-[¶]	78 (70-85)	-[¶]	-[¶]	-[¶]
Colditz	-	115	76	-	-	66 (57-74)	-	-	-
Teh	876	61	53	149	17.0	87 (76-93)	36 (28-44)	99 (97.7-99.3)	88.2 (85.8-90.3)
Machon	3355	176	39	48	0.1	22 (17-29)	81 (68-90)	99.6 (99.5-99.7)	99.9 (99.94-99.98)

Study	Study population[†] TP+FP+FN+TN	Self-report stroke TP+FP	Self-report confirmed TP	Total stroke TP+FN	Stroke prevalence (%) (TP+FN)/(TP+FP+FN+TN)	PPV (% & 95% CI) TP/(TP+FP)	Sensitivity (% & 95% CI) TP/(TP+FN)	Specificity (% & 95% CI) TN/(FP+TN)	NPV (% & 95%CI) TN/(TN+FN)
O'Mahony	1508	164	104	110	7.3	63 (56-70)	95 (89-98)	96 (94.5-96.7)	99.6 (99.0-99.8)
Heckbert	-	854	614	-	-	72 (69-75)	-	-	-
Okura	2,037	86[**]	58[**]	74[**]	3.6	67 (57-76)	78 (68-86)	98.6 (97.9-99.0)	98.6 (97.9-99.0)
Bergmann	-	113	76	-	-	67 (58-75)	-	-	-

[*] Table 3.1 Shows how PPV, sensitivity, specificity, and NPV were calculated.

[†]Unless otherwise stated, the study population is the number of participants (non responders excluded) for whom reference standard data was available. If left blank, the study only compared participants who self-reported stroke (TP+FP) against the reference standard.

[‡]These values were published as percentages.

[§]It was not possible to calculate sensitivity, specificity, or NPV in this study because validation was performed in a very small selected sample of the self-report negative ('no-stroke') participants.

[¶]In these studies, although the source of reference standard data was population based (GP questionnaire included to capture strokes diagnosed out of hospital), only those participants who self-reported stroke ('stroke positive' participants), plus a very small proportion of self-report negative ('no-stroke') were validated. There was insufficient data published to calculate sensitivity, specificity, NPV, or stroke prevalence.

[**]Calculated using published PPV, sensitivity, and number of strokes confirmed by medical record review.

Table 3.5 The proportion of self-reported strokes which were true stroke, true TIA, or either.

Study†	Country	Question asked‡	Self-report (n)	PPV True stroke (% & 95% CI)	PPV True TIA (% & 95% CI)	Revised PPV True stroke or TIA (% & 95% CI)
Machon	Spain	Stroke	176	22 (17–29)	6 (4–11)	28 (22–35)
Walker	UK	Stroke	201	56 (49–62)	25 (20–32)	81 (71–86)
O'Mahony	UK	Stroke/TIA	173	63 (56–70)	15 (11–22)	78 (71–84)
Engstad	Norway	Stroke	269	79 (74–84)	8 (4–10)	87 (82–90)
Bots	Netherlands	Stroke	285	67 (61–72)	10 (7–14)	77 (72–81)
Colditz	US	Stroke	115	66 (51–74)	21 (14–29)	87 (80–92)

TIA = Transient ischemic attack.
*According to the reference standard.
†Studies which published sufficient data (out of 17 included studies).
‡Participants were asked to report stroke, or stroke plus transient ischemic attack (TIA) (+/- synonyms for either).

doi:10.1371/journal.pone.0137538.t001

Stroke prevalence

PPV increased with increasing stroke prevalence (Figure 3.2).

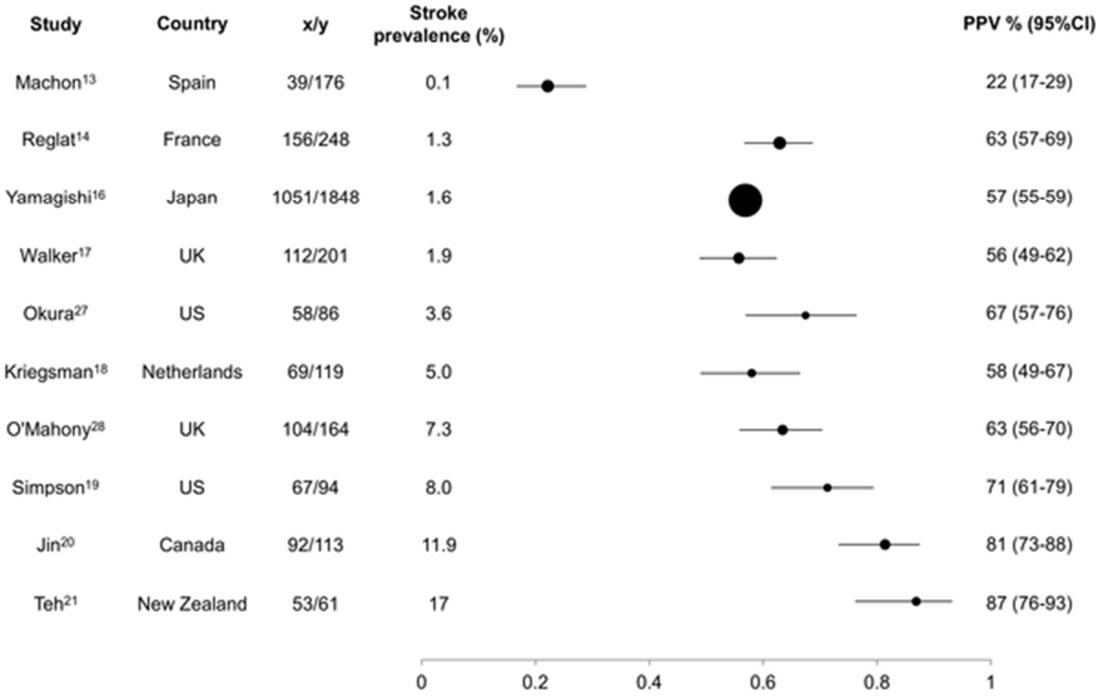

Figure 3.2 Influence of stroke prevalence on the PPV of self-report of stroke

x/y: x = number of self-reported strokes which are confirmed; y = total number of self-reported strokes

Participant age.

Among studies which reported the average age of responders (mean or median),(O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Engstad 2000, Jin et al. 2010, Okura et al. 2004, Barr et al. 2009, Teh et al. 2013) we noted increasing PPV

with increasing reporting age, probably because stroke prevalence increased with age (Table 3.6). Among five studies which published age at the time of self-report, and had sufficient data to calculate sensitivity,(O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Jin et al. 2010, Okura et al. 2004, Teh et al. 2013) the study with the highest mean participant age (84 years) had the lowest sensitivity for stroke (sensitivity 36%, 95% CI 28–44).(Teh et al. 2013) Sensitivity of self-report was stratified by age within one large study (~90,000 participants),(Yamagishi et al. 2009) and fell with increasing age (78% in those <75 years versus 69% in those ≥75 years). A similar pattern was observed in a second, smaller study (~1,536 participants. with sensitivities of 60% in those < 75 years versus 48% in those ≥ 75 years).(Jin et al. 2010) However, limited data for sensitivity as well as heterogeneity between studies in population characteristics meant that it was not possible to demonstrate a clear association between reporting age and sensitivity.

Table 3.6 The influence of age on PPV, Sensitivity, Specificity and NPV of self-report

Study*	Self-report (n)	Stroke prevalence (%)	Age† (mean)	PPV (% & 95% CI)	Sensitivity (% & 95% CI)	Specificity (% & 95% CI)	NPV (% & 95% CI)
Barr	87	-	52	38 (29–48)	-	-	-
Reglat	248	1.34	57	63 (57–69)	69 (62–74)	99 (99.3–99.6)	99.6 (99.5–99.7)
Engstad‡	269	-	60	79 (74–84)	-	-	-
Okura§	86	3.6	61¶	67 (57–76)	78 (68–86)	98.6 (97.9–99.0)	98.6 (97.9–99.0)
O'Mahony§	164	7	63	63 (56–70)	95 (89–98)	96 (94.5–96.7)	99.6 (99.0–99.8)
Jin	113	11.9	80	82 (73–88)	50 (43–57)	98 (97.6–98.9)	93.5 (92.1–94.7)
Teh	61	17	84	87 (76–93)	36 (28–44)	99 (97.7–99.3)	88.2 (85.8–90.3)

*Studies which published participant age at self-report (out of 17 included studies).
†Age at time of reporting
‡This study asked about stroke/synonyms
§These studies asked about stroke/TIA
¶Median

doi:10.1371/journal.pone.0137538.t002

Question(s) asked.

Overall, there was no clear difference in PPV or sensitivity between studies which asked about 'stroke' versus 'stroke plus synonyms' versus 'stroke/TIA'.(Walker et al. 1998, Britton et al. 2012, O'Mahony et al. 1995, Machon et al. 2013, Kriegsman et al. 1996, Engstad 2000, Yamagishi et al. 2009, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Heckbert et al. 2004, Bergmann et al. 1998, Teh et al. 2013)

However, among included studies there were no within study comparisons of the influence of the question(s) asked on PPV or sensitivity of stroke self-report.

Recall time.

In between study comparisons, recall of events over the last six months to one year (PPV 72% to 81%),(Jin et al. 2012, Heckbert et al. 2004) was not clearly more accurate than recall of events over the previous 2 to 5 years (PPV 38% to 78%),(Britton et al. 2012, Colditz et al. 1986, Barr et al. 2009) or over lifetime (PPV 22% to 87%).(Machon et al. 2013, Kriegsman et al. 1996, Fourrier-Reglat et al. 2010, Engstad 2000, Bots et al. 1996, Simpson et al. 2004, Okura et al. 2004, Teh et al. 2013) Two studies stratified PPV results by recall time, and neither found a difference in PPV for more versus less recent events.(Yamagishi et al. 2009, Barr et al. 2009) One of these studies (~ 90,000 participants) found no difference in sensitivity for more versus less recent events.(Yamagishi et al. 2009

3.5 Discussion

As far as I am are aware this is the first systematic review of the accuracy of patient self-report of stroke. Self-report had variable PPV (range 22 to 87%) and variable sensitivity (range 36 to 98%) for stroke, but consistently high specificity and NPV (88 to 99%). In populations with low stroke prevalence, it would take a very large number of false positives to reduce specificity or NPV. PPV and sensitivity are therefore more informative measures. PPV increased with increasing stroke prevalence. Although this relationship is not surprising, we have shown that in populations with low stroke prevalence (<10%), a large proportion of self-reported strokes (~1/3 to 3/4) were false positive. This has important implications for large prospective studies, where stroke prevalence is likely to be low.

Reviews of the accuracy of self-report of various diseases have found that PPV and sensitivity vary depending on the disease reported.(Harlow and Linet 1989) For certain diseases, such as myocardial infarction, or cancer, a large proportion of false positive self-reports occur because patients confuse the diagnosis with a similar condition (e.g. other cardiovascular disease, or other cancer type).(Paganini-Hill and Chao 1993) Similarly, we found that 6–25% of individuals who self-reported stroke had a reference standard diagnosis of TIA. If doctors or other health professionals used the term ‘mini-stroke’ when referring to TIA, the patient may be misled into thinking they had had a stroke. Grouping of stroke and TIA cases might be acceptable for some research questions (e.g. those which explore common risk factors for stroke and TIA combined). If both TIA and stroke were considered true positive, the PPV of self-report of stroke (or stroke/TIA) was higher.

Other research questions require accurate identification of stroke, and accurate exclusion of TIA cases, (e.g. those which explore risk-factors associated with the different pathological types and subtypes of stroke). However, there is no ‘gold standard’ diagnosis for stroke. The classic ‘symptom-based’ definition of stroke relies on symptom duration (>24 hours) to distinguish stroke from TIA.(Hatano 1976) A newer ‘tissue-based’ definition has been proposed, which relies on the presence of brain infarction to diagnose stroke, irrespective of symptom duration (<24 hours),(Albers et al. 2002) but application of this rule reclassifies cases of ‘TIA

with infarction' as stroke. Although physicians (expert and non expert) are inconsistent in diagnosing stroke using 'symptom-based' definitions,(Ferro et al. 1996b) the 'tissue-based' definition is equally susceptible to inter-observer variation.(Fiebach et al. 2002, Wardlaw and Mielke 2005, Brown et al. 2003a) Accurate diagnosis of brain infarction depends on the choice and timing of imaging, and on reviewer expertise.(Fiebach et al. 2002, Wardlaw and Mielke 2005) The 'tissue-based' definition is therefore likely to be particularly susceptible to variation when applied across different settings (with different brain imaging resources).(Brown et al. 2003a) To maintain consistency, we excluded studies which used the 'tissue-based' definition from our review. However, as new definitions and diagnostic terms continue to emerge, this lack of consistency will contribute to the misreporting of stroke (as TIA and vice versa) by patients and their physicians. Previous primary studies have assessed the influence of gender, cognitive impairment, education, and number of co-morbidities on the accuracy or reliability of stroke self-report, with variable and sometimes conflicting results.(Kriegsman et al. 1996, Engstad 2000, Simpson et al. 2004, Jin et al. 2010, Okura et al. 2004, Bergmann et al. 1998, Teh et al. 2013) However, it is difficult to draw overall conclusions because a range of different methods were used to analyse data and present results.

We observed a wide range in PPV and sensitivity of self-report, which is likely to reflect between study heterogeneity in both population characteristics and study design. Reassuringly, only a few studies had low response rates (<60%). While this may have introduced selection bias, the accuracy (range of PPVs) of self-report was not clearly affected by response rates. Neither was there any clear effect of incomplete blinding on the range of PPVs. The majority of included studies had missing reference standard data.(Walker et al. 1998, O'Mahony et al. 1995, Fourrier-Reglat et al. 2010, Kriegsman et al. 2006, Engstad 2000, Bots et al. 1996, Simpson et al. 2004, Jin et al. 2010, Colditz et al. 1986, Bergmann et al. 1998, Barr et al. 2009, Teh et al. 2013) Although this is an important source of potential bias, incomplete reporting meant that it was not possible to assess the influence of missing reference standard data on the PPV and/or sensitivity of self-report. The reference standard used (hospital versus population based) was an additional potential source of bias.

Studies which excluded cases diagnosed out of hospital from their reference standard had a higher chance of misclassifying ‘true stroke cases’ as ‘false positive’ reports, and so of underestimating PPV. However, we did not find a difference in the overall accuracy of self-report (PPV) between hospital based and population based studies.

Strengths of this study include our thorough search strategy, adherence to published guidelines for test accuracy reviews,(deVet et al. 2008) and inclusion of all relevant studies of stroke self-report. Although we only searched two online databases, a strategy which may have missed potentially relevant articles, we augmented our search by screening bibliographies of all included publications. Bibliography screening may be the most effective method of identifying additional relevant articles in systematic reviews of test accuracy.(Whiting et al. 2008) Additional strengths of our review include the exclusion of studies which failed to distinguish TIA from stroke in the reference standard, use of a single stroke definition (WHO),(Hatano 1976) and exclusion of studies which used coded data as the only source of stroke confirmation.

There were some limitations. First, variation in the accuracy and completeness of the reference standard may have contributed to between study variation in PPV and sensitivity. To improve comparability between studies, we only calculated sensitivity, specificity, NPV, or stroke prevalence when the reference standard was population based. This was possible in ten out of seventeen included studies. However, only two included studies used the most robust population based reference standard for stroke,(Sudlow and Warlow 1996) with multiple sources of case ascertainment and confirmation.(Britton et al. 2012, O’Mahony et al. 1995) Second, the true sensitivity of self-report is likely to be lower than the included studies suggest, since non responders could not be included in the denominator population (non response ranged from 10–77% amongst included studies). Third, incomplete reporting and limited within study comparisons of population characteristics (such as age, gender, education, cognitive impairment, comorbidities) made it difficult to assess the influence of these individual factors on self-report PPV or sensitivity. Fourth, as discussed above, there is no gold standard test to diagnose stroke or TIA.

Lack of consistency in determining the 'true' diagnosis is likely to have contributed to the wide range of reported PPV, sensitivity and stroke prevalence.

Further work is needed to assess and compare multiple overlapping sources of stroke detection in large epidemiological studies. Some studies have found that self-report increases the number of potential strokes detected (compared to hospital or primary care data).(Britton et al. 2012, Heckbert et al. 2004) However, it is uncertain whether using self-report is time- or cost-effective for stroke case detection in large prospective studies, because potential strokes would need to be confirmed, for example by medical record review. In addition, future work should examine the influence of the question asked on PPV and/or sensitivity of stroke report. We did not find a clear influence of the questions asked on PPV or sensitivity, but there were no within study comparisons of stroke specific questions versus stroke/TIA or stroke synonyms. Establishing the best list of questions (to improve disease specific sensitivity or PPV) will be important for future questionnaire design. Future work could also consider the influence of new stroke definitions (where used), and more sensitive imaging methods (where available) on the PPV and/or sensitivity of patient self-report.

3.6 Conclusions

Based on the results of this study, I suggest that self-report of stroke may be a useful screening tool to identify potential stroke disease in prospective studies, but not accurate enough on its own to confirm cases. Once potential cases are identified, a subsequent confirmation step using other data sources will also be required. The influence of stroke prevalence on PPV means that in studies with low stroke prevalence, like UK Biobank, a large proportion of potential strokes identified by self-report may be false positives.

3.7 Appendix

3.7.1 Review Question(s)

Primary Question

Accuracy (Positive Predictive Value) of patient self-report of stroke (for example by self administered questionnaire, or face-to-face interview) for a diagnosis of stroke (using WHO or equivalent definitions) in an adult population.

Secondary Questions

A. Sensitivity and specificity of patient self-report of stroke (amongst studies where the reference standard is population-based).

B. Influence of stroke prevalence on the accuracy (PPV) of self-report of stroke (amongst studies where the reference standard is population based).

C. Accuracy (Positive Predictive Value) of patient self-report of stroke for a diagnosis of Transient Ischaemic Attack (TIA).

D. Influence of the question(s) asked (for example stroke, or stroke plus TIA, or stroke plus TIA/synonyms for either) on the accuracy (PPV, sensitivity, specificity) of self-report of stroke.

E. Influence of period of recall (for example lifetime history versus more recent events) on the accuracy (PPV, sensitivity, specificity) of self-report of stroke.

F. Influence of participant age (mean, years) on the accuracy (PPV, sensitivity, specificity) of self-report of stroke.

3.7.2 Searches

We will search the following databases from inception to the date of search:

MEDLINE (Ovid SP);

EMBASE (Ovid SP)

Cochrane Register of Diagnostic Test Accuracy Studies

We will review bibliographies of included publications for any additional relevant articles.

MEDLINE search strategy

1. cerebrovascular disorders/ or exp basal ganglia cerebrovascular disease/ or exp brain ischemia/ or exp carotid artery diseases/ or exp cerebral small vessel diseases/ or exp intracranial arterial diseases/ or exp "intracranial embolism and thrombosis"/ or exp intracranial hemorrhages/ or stroke/ or exp brain infarction/ or stroke, lacunar/ or vertebral artery dissection/

2. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or apoplex$ or isch?emi$ attack$ or tia$ or SAH).tw.

3. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

4. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraventricular or infratentorial or supratentorial or basal gangli$ or subarachnoid or putaminal or putamen or posterior fossa) adj5 (haemorrhage$ or hemorrhage$ or haematoma$ or hematoma$ or bleed$)).tw.

5. 1 or 2 or 3 or 4

6. questionnaires/ or self report/

7. self concept/ or self-assessment/ or self disclosure/ or diagnostic self evaluation/

8. Interviews as Topic/

9. Medical History Taking/

10. truth disclosure/

11. (self-report$ or questionnaire$).tw.

12. (patient$ adj5 report$).tw.

13. 6 or 7 or 8 or 9 or 10 or 11

14. "reproducibility of results"/ or "sensitivity and specificity"/ or "predictive value of tests"/

15. (positive predictive value or sensitivity or specificity).tw.

16. (agreement or validity or reliability or reproducibility or accuracy or accurate or concordance).tw.

17. 14 or 15 or 16

18. 5 and 13 and 17

19. records as topic/ or hospital records/ or exp medical records/ or nursing records/

20. exp Diagnosis/

21. ((diagnosis or diagnosed or history) adj5 stroke).tw.

22. ((hospital or GP or medical or general practitioner or health) adj5 (record or records)).tw.

23. cerebrovascular disorders/di or exp basal ganglia cerebrovascular disease/di or exp brain ischemia/di or exp carotid artery diseases/di or exp cerebral small vessel diseases/di or exp intracranial arterial diseases/di or exp "intracranial embolism and thrombosis"/di or exp intracranial hemorrhages/di or stroke/di or exp brain infarction/di or stroke, lacunar/di or vertebral artery dissection/di

24. 19 or 20 or 21 or 22 or 23

25. 18 and 24

EMBASE search strategy

1. cerebrovascular disease/ or basal ganglion hemorrhage/ or exp brain hematoma/ or exp brain hemorrhage/ or exp brain infarction/ or exp brain ischemia/ or exp carotid artery disease/ or cerebral artery disease/ or cerebrovascular accident/ or exp occlusive cerebrovascular disease/ or exp stroke/

2. stroke unit/ or stroke patient/

3. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or apoplex$ or isch?emi$ attack$ or tia$ or SAH).tw.

4. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

5. ((brain$ or cerebr$ or cerebell$ or intracerebral or intracran$ or parenchymal or intraventricular or infratentorial or supratentorial or basal gangli$ or subarachnoid or putaminal or putamen or posterior fossa) adj5 (haemorrhage$ or hemorrhage$ or haematoma$ or hematoma$ or bleed$)).tw.

6. 1 or 2 or 3 or 4 or 5

7. self-report/ or exp questionnaire/ or exp interview/

8. self concept/ or self evaluation/ or self disclosure/

9. health assessment questionnaire/ or health perceptions questionnaire/ or interpersonal communication/

10. anamnesis/ or medical history/

11. (self-report$ or questionnaire$).tw.

12. (patient$ adj5 report$).tw.

13. 7 or 8 or 9 or 10 or 11 or 12

14. "sensitivity and specificity"/

15. exp validity/ or exp reliability/ or reproducability/ or accuracy/ or predictive value/

16. (positive predictive value or sensitivity or specificity).tw.

17. (agreement or validity or reliability or reproducibility or accuracy or accurate or concordance).tw.

18. 14 or 15 or 16 or 17

19. 6 and 13 and 18

20. medical record/ or medical record review/ or electronic medical record/

21. exp diagnosis/

22. ((diagnosis or diagnosed or history) adj5 stroke).tw.

23. cerebrovascular disease/di or basal ganglion hemorrhage/di or exp brain hematoma/di or exp brain hemorrhage/di or exp brain infarction/di or exp brain ischemia/di or exp carotid artery disease/di or cerebral artery disease/di or cerebrovascular accident/di or exp occlusive cerebrovascular disease/di or exp stroke/di

24. ((hospital or GP or medical or general practitioner or health) adj5 (record or records)).tw.

25. 20 or 21 or 22 or 23 or 24

26. 19 and 25

3.7.3 Types of study to be included

We will include studies in any adult population (cohort studies, case control studies, or clinical trials) which compare the self-report of stroke against a reference standard diagnosis of stroke (WHO, or equivalent definitions).

Studies are required to report the Positive Predictive Value of participant self-report (or data from which this can be calculated).

Studies should use a reference standard of ‘stroke’ (distinguished from transient ischaemic attack or generalised cerebrovascular disease) when calculating PPV, sensitivity, and specificity.

We will exclude studies which assess < 50 self-reported strokes (due to limited precision).

3.7.4 Condition or domain being studied

Index test: Self-report of stroke, or ‘stroke plus TIA’, or ‘stroke plus TIA/synonyms for either’ by any method (questionnaire or interview).

Studies should include the term ‘stroke’ in their questionnaire or interview. This excludes studies which assess self-report of symptoms, ‘cerebrovascular disease’, or past medical history for stroke.

3.7.5 Participants/population

Any adult population.

We will not exclude studies based on participant selection criteria (eg. age, cardiovascular risk, education, cognitive impairment, disability).

We have specified secondary questions which will examine the influence of participant characteristics (eg. age and stroke prevalence) on the accuracy of self-report.

3.7.6 Interventions, exposures

Reference standard: We accept that the inter-observer reliability of stroke diagnosis is imperfect, even amongst experts. In the absence of a true ‘gold standard’ for stroke, we will include studies which use any of the following reference standards for stroke: clinical examination; physician questionnaire; medical record review

(primary care and/or hospital records); stroke registers (informed by multiple overlapping data sources, 'hot pursuit', and expert medical record review).

To improve accuracy we will exclude studies which use coded data (eg. International Classification of Diseases codes) for the reference standard, unless other methods (above) are also used.

Studies should use a symptom-based definition (WHO, or equivalent) for diagnosing stroke.

3.7.7 Comparators/control

Not applicable

3.7.8 Context

The principle criterion is self-report of stroke. We aim to inform approaches to stroke ascertainment in large population based studies.

Depending on the design of the study, the measure of interest could be self-report of prevalent (lifetime stroke) or incident (since recruitment) events. We have pre specified questions to see if the accuracy of self-report varies by length of recall time.

In UK Biobank, participants were asked to self-report stroke by questionnaire, and answers were later confirmed during a brief nurse led interview. We will include studies which assess self-report by either self administered questionnaire or face-to-face interview.

3.7.9 Outcomes

Primary outcomes

We will calculate Positive Predictive Value (PPV) of self-report for stroke in all included studies using the available published data.

Secondary outcomes

In studies which use population-based reference standards, we will calculate Sensitivity, Specificity, PPV, and stroke prevalence using 2x2 contingency tables.

The reference standard will be grouped into hospital based versus population based according to the following definitions:

Population based: a reference standard which identifies strokes diagnosed out of hospital, and therefore captures as many 'true' strokes in the participant population as possible; primary care medical records; general practitioner questionnaires; population-based stroke registers (informed by expert medical record review); physician assessment of all participants.

Hospital based: a reference standard which only identifies hospitalised strokes; hospital medical records; hospital physician questionnaires; hospital based stroke registers (informed by expert hospital record review).

3.7.10 Data extraction (selection and coding)

We will extract data onto study specific proforma.

Covariates of interest:

- Study author
- Country
- Population selection criteria (eg. random or consecutive sampling, inclusions/exclusion criteria)
- Self-report method (postal questionnaire, telephone interview, face-to-face interview)
- Question(s) asked
- Participant response rates (% population who agree to take part, complete questionnaires, and/or attend interviews).
- Participant age at self-report (mean, median or range).
- Length of recall time (years, or lifetime)
- Reference Standard used (eg., medical records, physician questionnaire and whether hospital or population based)
- Number(s) included in final analysis (2x2 table, or true positive versus false positive self-reported strokes for PPV calculation).

- Number(s) excluded from analysis (eg. because of incomplete reference standard data).
- Blinding (to self-report status for determination of the reference standard diagnosis)

3.7.11 Risk of bias (quality) assessment

We will assess methodological quality using the revised Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2). We will focus on assessment of bias questions rather than generalizability.

Modified QUADAS-2 questions for assessment of bias:

1) Patient selection

a) Sampling method: was consecutive or random sampling used?

Low risk of bias: consecutive or random sampling was used. Unclear risk of bias: insufficient information published. High risk of bias: did not use consecutive or random sampling.

b) Study design: was a case-control study design avoided?

Low risk of bias: case-control design avoided. High risk of bias: case-control design used. Unclear risk of bias: insufficient information published.

c) Population included: were inappropriate exclusions avoided?

Low risk of bias: inappropriate exclusions were avoided. High-risk of bias: participants were excluded based on characteristics which might influence self-report accuracy (eg, education level, cognitive impairment, disability.) Unclear risk of bias: insufficient information published.

2) Index test (self-report)

a) Interpreted blind: was the index test interpreted without knowledge of the reference standard diagnosis?

Risk of bias depends on the degree of interpretation of the index test (self-report) and the presence/absence of blinding to the reference standard diagnosis. Low risk; a ‘yes/’no’ answer given by the participant in a self-administered questionnaire, irrespective of blinding; self-report by face-to-face interview and blinding present. Unclear risk: more than a ‘yes’/’no’ answer in a self-administered questionnaire and blinding not reported; a ‘yes’/’no’ answer in a face-to-face interview and blinding not reported; question(s) asked not published. High risk: more than a ‘yes’/’no’ answer in a face-to-face interview and blinding either not reported or not present;

more than a ‘yes’/’no’ answer in a self-administered questionnaire and blinding not present; face-to-face interview by medically trained professional, irrespective of blinding.

b) Was the threshold pre-specified?

Selecting the test threshold to optimize sensitivity/specificity may lead to overestimation of test performance. Low risk of bias: the question(s) asked were published and the answers accepted as positive self-report (e.g., stroke/stroke plus TIA/stroke plus TIA plus synonyms for either) were specified at the outset. Unclear risk of bias: the question(s) asked were not published but it was specified at the outset which answers were accepted as positive self-report. High risk of bias: the question(s) asked were not published and it was not specified at the outset which answers were accepted as positive self-report.

3) Reference standard

a) Source of data: is the reference standard likely to correctly classify the target condition?

Low risk of bias: the reference standard was ‘population-based’ and included ‘true stroke’ cases diagnosed out of hospital. Unclear risk of bias: insufficient information. High risk of bias: the reference-standard was ‘hospital-based’ and excluded ‘true stroke’ cases diagnosed out of hospital.

b) Blind to self-report: were reference standard results interpreted without knowledge of results of the index test?

Low risk of bias: the reference standard diagnosis was made blind to self-report results. Unclear risk of bias: blinding not reported. High risk of bias: the reference standard diagnosis was made with knowledge of self-report results.

4) Flow and timing

a) Participant response rates: were all patients included in the analysis?

Low risk of bias: participant response rates >80%. High risk of bias: participant response rates <80%. Unclear risk of bias: participant response rates not reported.

b) Missing data: did all patients receive a reference standard?

Low risk of bias: reference standard data available for all responding participants. Unclear risk of bias: insufficient information published. High risk of bias:

responding participants excluded from the final analysis due to missing reference standard data or because the final reference standard diagnosis was ‘unclear’.

c) Differential verification: did all patients receive the same reference standard?

Low risk of bias: the same reference standard was used for all participants. Unclear risk of bias: different reference standards were used but it was unclear if this differed by self-report status. High risk of bias: a different reference standard was used if participants self-reported stroke (vs. self-reported no-stroke).

d) Was there an appropriate interval between the index test and the reference standard?

This question was excluded from our study because the time interval between the index test (self-report) and reference standard (‘true’ diagnosis of stroke versus non-stroke) did not influence the accuracy of the original diagnosis (which was made historically). The time interval could influence the availability of reference standard data, but this aspect of bias was already captured in question 4b (above).

3.7.12 Strategy for data synthesis

We will cross classify self-reported strokes with the reference standard diagnosis (‘stroke’ versus ‘non-stroke’).

PPV (%) = true positive reports / [true positive reports + false positive reports] x 100

True positive reports = self-report ‘stroke’ and reference standard ‘stroke’.

False positive reports = self-report ‘stroke’ and reference standard ‘non-stroke’.

Where the reference standard is population based, we will construct standard 2x2 tables describing binary test results (self-report ‘stroke’ and self-report ‘non-stroke’) cross classified with binary reference standard results (‘stroke’ and ‘non-stroke’).

We will use this data to calculate sensitivity, specificity, PPV, and 95% confidence intervals. Stroke prevalence will be calculated using ‘total reference standard stroke’/ ‘final included population’.

We will tabulate results for visual inspection to assess the influence of participant age, question(s) asked, recall time, and stroke prevalence on the PPV, sensitivity

and/or specificity of self-report. Where possible, and to limit the impact of between-study heterogeneity, we will use within-study as well as between-study comparisons.

We will assess heterogeneity between studies by inspection of tabulated data.

We will not quantify publication bias as there is no assessment applicable to test accuracy.

3.7.13 Dissemination plans

We will present our findings at local, national and international meetings. We plan to publish a full paper in a peer-reviewed scientific journal.

Chapter 4 Reliability and feasibility of ischaemic stroke classification systems for large epidemiological studies: a systematic review

- Ischaemic stroke subtypes may differ in the strength of their risk factor associations. Very large prospective studies, with large numbers of anticipated stroke outcomes, are required to reliably test these associations.
- Ischaemic stroke classification systems are ‘anatomical’ or ‘mechanistic’.
- An ideal ischaemic stroke classification system for UK Biobank would classify large numbers of ischaemic stroke cases to a single underlying mechanism without compromising reliability. It would be scalable and applicable in different clinical settings.
- In this chapter I report my systematic review of the reliability of existing ischaemic stroke classification systems.
- I used a comprehensive search strategy, critically appraised study quality, assessed factors which affected reliability, and reported the proportion of cases which were assigned to ‘undetermined cause’.
- Reliability of classification varied widely, reflecting heterogeneity of study settings as well as the classification system(s) used.
- I found that clear rules, data abstraction protocols, computer based classification and fewer categories all improved reliability.
- There was insufficient evidence to suggest that newer ischaemic stroke classification systems were superior to older, more established systems.
- I conclude that no single system is fit for every purpose, and recommend multiple approaches to ischaemic stroke classification in UK Biobank.

4.1 Introduction

The main types and subtypes of stroke are complex clinical conditions with multiple shared common risk factors. However, evidence is increasingly emerging that these distinct conditions are likely to differ in the strength of their risk factor associations

and in their causal pathways. (Jackson et al. 2010, Jackson and Sudlow 2006, Doubal et al. 2009, O'Donnell et al. 2010, Asia Pacific Cohort Studies Collaboration 2005, Biffi et al. 2010) Detecting potential differences in the strength of associations between shared common risk factors, such as blood pressure, and the risk of stroke types and sub-types requires very large observational studies, often with thousands of clinical outcomes for adequate statistical power.(Burton et al. 2009) The different outcomes also need to be classified accurately because misclassification reduces power to detect potential differences in the strength of associations between them.(Jaffar et al. 2003b, Choi et al. 2002b) UK Biobank offers the perfect opportunity to explore the complex determinants of the main types and subtypes of stroke because of the detailed baseline assessment (collecting accurate information on a wide range of potential exposures and covariates), the large numbers of anticipated cases, and the planned detailed process of stroke detection, confirmation and sub-classification.

In this chapter I focus on exploring the best potential method of classifying confirmed ischaemic stroke cases into their main pathological subtypes.

4.1.1 Existing ischaemic stroke classification systems

In clinical practice, ischaemic stroke classification depends largely on an individual physician's judgement rather than clearly defined rules, which is likely to generate variability and possibly systematic biases. Observational studies and clinical trials have developed rule based systems to improve the reliability of ischaemic stroke classification. The main approaches use either the presenting clinical syndrome to predict infarct site and size; 'anatomical classification', or use the presenting clinical syndrome plus results of investigations to identify the underlying pathological mechanism of infarction; 'mechanistic classification'. Figure 4.1 displays the main approaches to ischaemic stroke classification.

The Oxford Community Stroke Project (OCSP) is the main 'anatomical' method of ischaemic stroke classification (Bamford et al. 1991) It predicts the vascular territory of infarction with up to 75% accuracy based on clinical features at the time of presentation.(Mead et al. 2000) The four main subtypes of OCSP classification are

Total Anterior Circulation Stroke (TACS), Partial Anterior Circulation Stroke (PACS), Posterior Circulation Stroke (POCS), and Lacunar Stroke (LACS). Brain imaging is used to enhance accuracy, but is not an absolute requirement for classification.(Potter et al. 2010) When brain imaging is used, OCSP reliably distinguishes 'small vessel disease' strokes (LACS) from 'large vessel disease' strokes (TACS, PACS, POCS). Although it distinguishes small vessel disease from large vessel disease, it does not identify the main mechanisms which cause large vessel ischaemic stroke, for example cardiac embolism or large artery atherosclerosis.

Mechanistic classification systems can be either 'single cause', assigning the single most likely stroke mechanism, or 'descriptive', describing the contribution of each mechanism to the stroke, which means that multiple different mechanisms can be considered and, if felt appropriate, can be classified with equal contribution to the causative pathway (see Figure 4.1).

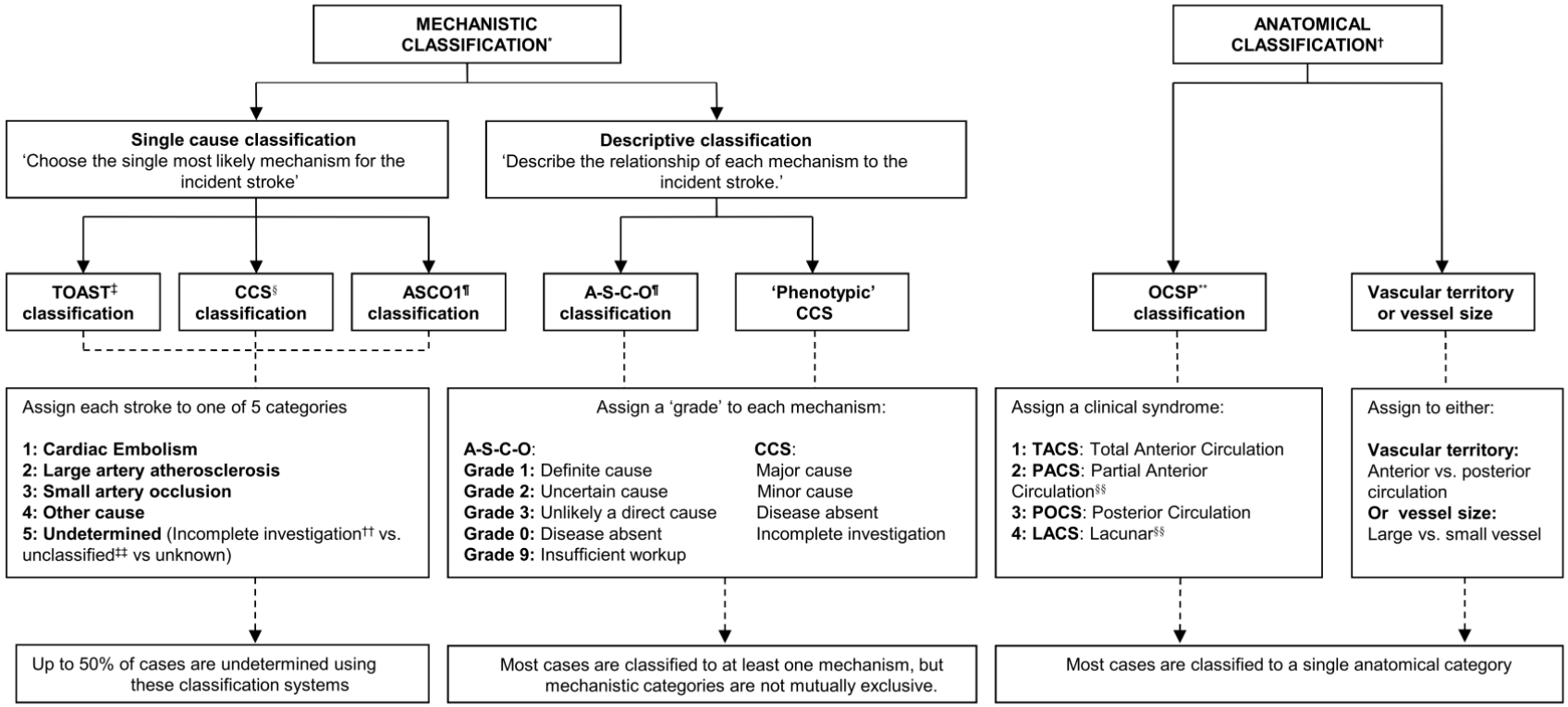

Figure 4.1 The main approaches to ischaemic stroke classification

*Mechanistic classification requires investigation of the three main mechanisms of ischaemic stroke: Large artery atherosclerosis (LAA), Cardiac Embolism (CE), and Small artery Occlusion (SAO).

†Anatomical classification is based on the presenting clinical syndrome, but brain imaging may be used to improve accuracy.

‡Trial of Org 10172 in Acute Stroke Treatment classification.

[§]Causative Classification System.

[¶]The single cause mechanistic system ‘ASCO1’ is derived from the descriptive classification system ‘A-S-C-O’: Atherosclerosis, Small vessel disease, Cardiac source, Other cause classification. ‘ASCO1’ classifies to a ‘single cause’ when a single mechanism has ‘grade 1’ evidence (eg A1, S1, C1 or O1). Cases are ‘undetermined’ by ASCO1 if there is more than one ‘grade 1’ mechanism, no ‘grade 1’ mechanism, or any ‘grade 9’ mechanism.

[**]Oxfordshire Community Stroke Project classification.

[††]Complete investigation requires that minimum investigations have been performed to evaluate every mechanism routine blood tests, brain imaging, cardiac investigations, evaluation of the extra +/- intracranial blood vessels. Additional investigations may be required if other rare causes (Oth) are considered likely.

[‡‡]Due to multiple potential mechanisms.

[§§]Distinction between LACS and PACS is up to 20% inaccurate when based on the clinical syndrome alone.

The TOAST (Trial of Org 10172 in Acute Stroke Treatment) system is the most established 'single cause' mechanistic method. It classifies infarcts into one of four determined categories: Large Artery Atherosclerosis (LAA), Cardiac Embolism (CE), Small artery occlusion (SAO), or Other (Oth) when clinical presentation, imaging and ancillary investigations point to a single underlying cause. If two or more potential mechanisms are present, cases are assigned to an 'undetermined' category: 'unclassified due to multiple potential mechanisms'.(Adams et al. 1993) The total proportion of cases undetermined by single cause mechanistic classification systems is the proportion unclassified due to multiple potential mechanisms, incompletely investigated, or unknown despite complete investigation. Up to 50% of cases are 'undetermined' by TOAST.(Ay 2010) A large 'undetermined' proportion is a significant disadvantage because these cases are often excluded from research studies. Exclusion of large numbers of outcomes limits statistical power and, depending on the research question, may introduce case selection bias.

The Causative Classification System (CCS)(Ay et al. 2005a) is a more recently developed 'mechanistic' classification system. This modified version of TOAST uses a hierarchy of evidence to decide between 'multiple potential mechanisms' where they exist, and assign a single mechanism. It aims to reduce the undetermined proportion, compared to TOAST, without compromising the reliability of classification.(Ay et al. 2007)

A third major 'single cause mechanistic system' is ASCO1 (Atherosclerosis, Small Artery Disease, Cardiac Embolism, Other-1). ASCO1 is derived from A-S-C-O (a descriptive mechanistic system).(Amarenco et al. 2009b) A-S-C-O considers each potential mechanism of ischaemic stroke separately. It assigns a score to each mechanism, Atherosclerosis (A), Small artery disease (S), Cardiac embolism (C), or Other (O), depending on the presence or absence of disease and the contribution of that disease to the incident stroke (1 for potentially causal, 2 for causality is uncertain, 3 for unlikely causal but the disease is present, 0 for absence of disease, and 9 for insufficient workup to rule out the disease). ASCO1 is the 'single cause mechanistic' counterpart of A-S-C-O. ASCO1 classifies to a 'single cause' when a single mechanism has scored level 1 evidence for 'potentially causal'

(eg. A1, S2-3/0/9, C2-3/0/9, or O2-3/0/9). Cases are 'undetermined unclassified' by ASCO1 if there are multiple potentially causal mechanisms present (eg., A1 and S1/C1/O1), 'undetermined unknown' if there is no potentially causal mechanism (eg., score 2, 3 or 0 for all mechanisms), or 'undetermined-incomplete investigation', if any mechanism is incompletely investigated ('score 9' for any mechanism).

The main descriptive mechanistic systems are A-S-C-O (as described above), and phenoyptic CCS. This version of CCS describes the contribution of each potential mechanism (large artery disease, cardiac embolism, small artery occlusion, other) to the incident stroke, described as: major disease; minor disease; absent disease; incompletely investigated.(Arsava et al. 2010) These systems have a very large number of descriptive categories (e.g. using A-S-C-O, a single case could be classified to any one of 625 potential categories).

4.1.2 Reliability and agreement

Reliability and agreement (terms often used interchangeably) are ways of demonstrating reproducibility: the degree to which repeated measurements in stable study objects provide similar results. Reliability measures how well objects can be distinguished from each other despite measurement errors, while agreement assesses exactly how close the scores for repeated measurements are.(de Vet et al. 2006) There are various measures of reliability depending on the data format and nature: the kappa statistic (measure of 'true agreement' i.e. proportion of agreement beyond that expected by chance) used for nominal and ordinal data; ranked intra-class correlation used for ordinal data; and intra class correlation coefficient used for continuous data.(Kottner et al. 2011) Results of reliability and agreement studies provide information about the amount of error inherent in a classification which in turn may affect its validity.(Kottner et al. 2011).

4.1.3 Aims

In this chapter I report a systematic review of published studies of ischaemic stroke classification systems, with the aim of identifying a feasible approach for ischaemic stroke classification in UK Biobank. An ideal classification system for large scale use would assign the maximum number of cases to a single determined subtype

without sacrificing reliability or positive predictive value, and would be robust to variations in the investigations performed in different clinical settings. It would perform well retrospectively, using clinical data routinely available in the UK, and the results would be reproducible between multiple observers across multiple centres.

My focus here is on the inter-observer reliability of ischemic stroke classification systems and the factors which affect these measures. I assess the inter- and/or the intra-observer reliability of ischaemic stroke classification systems in any human adult population with confirmed ischaemic stroke. Due to the importance of the undetermined proportion and the development of newer classification systems which attempt to reduce this, I also compare the proportion of undetermined cases amongst included studies.

4.2 Methods

4.2.1 Search strategy

I searched Ovid Medline and Embase (1990 to November 2013) for studies which assessed inter- or intra-observer reliability of ischaemic stroke classification systems in human adults (see Appendix 4.6.1). I included foreign language articles, obtaining translations were necessary. I assessed eligibility by reviewing all titles and abstracts, and the full text of potentially relevant articles, and resolved uncertainties through discussion with my supervisor, Professor Cathie Sudlow.

4.2.2 Eligibility criteria

I included studies which reported inter- and/or intra-observer reliability (kappa statistic) of an anatomical or mechanistic ischaemic stroke classification system. Studies had to: describe or reference the classification system; use CT or MR brain imaging to exclude intracerebral haemorrhage in the majority of cases (>80%); and report the proportion of cases excluded from analyses. I excluded studies published before 1990 (since clinical practice and classification systems have evolved substantially since then), duplicate studies (e.g. data published in abstract form if later published in full), studies assessing classification of ischaemic strokes in children (<18 years), or studies assessing classification systems in highly selected cases (e.g. those involving a particular vascular territory only). I excluded studies which reported the proportion of cases classified to mutually exclusive subtypes if they did not also report observer reliability (as inter/intra-observer reliability was the focus of my review). I accepted anatomical definitions if they referred to the vascular territory of an infarct and encompassed both anterior and posterior circulations. Mechanistic systems had to include large artery atherosclerotic (LAA), small artery occlusion (SAO) and cardio-aortic or cardio-embolic (CE) subtypes.

4.2.3 Data extraction and synthesis

I extracted information from included studies on the classification system(s) studied, number of cases classified, inclusion/exclusion of haemorrhagic strokes and stroke mimics, number of ratings made, inter- and intra-observer reliability, and the proportion of cases assigned to an undetermined subtype. I extracted both the overall kappa statistic for reliability across all potential categories, and subtype kappa

statistics. For most classification systems, the subtype kappa measures the reliability of classifying one subtype versus any other, while for descriptive mechanistic systems it measures the reliability of grading the contribution of each separate mechanism to every case (Figure 4.1). The proportion of cases undetermined by single cause mechanistic classification systems was the proportion unclassified due to multiple potential mechanisms, incompletely investigated, or unknown despite complete investigation.

I assessed potential risk of bias using criteria developed specifically for this study that I considered might influence reliability results (presence/absence of blinding between observers, and involvement or not of authors or adjudicators in classification system development).

Informed by the established Guidelines for Reporting Reliability and Agreement Studies (GRRAS),(Kottner et al. 2011) I summarised additional characteristics which might affect reliability including: population characteristics (country, population sampled, age range, case selection methods, time since stroke); observer characteristics (training in classification methods, single versus consensus observations, single versus multiple institutions, similar versus different expertise); data presented to adjudicators (clinical examination, original medical records, abstracted data, or a combination of sources), methods for assigning categories (number of categories used, use of data abstraction protocols and/or computer based assignment); and time between repeated observations.

I displayed kappa statistics in tables and/or forest plots stratified by the classification system used, and, where possible, by characteristics (above) which might affect reliability. I also examined the influence of individual investigations performed and the level of detail of investigation on reliability and the proportion of cases allocated to an undetermined category. For all reliability comparisons, I focused on within study comparisons rather than using a meta-regression approach, due to substantial heterogeneity in all study characteristics, and because the required data for meta-analysis were unavailable for a substantial proportion of included studies.

4.3 Results

From 1,105 publications identified, 29 studies were eventually included (Figure 4.2).

4.3.1 Study characteristics

25 studies assessed the overall inter-observer reliability of one or more ischaemic stroke classification systems, three assessed subtype inter-observer reliability only,(Marnane et al. 2010, Chen, Zhou and Wang 2013, Johnson et al. 1995) and one assessed intra-observer reliability only.(Cotter, Belham and Martin 2012) Studies were performed in Europe (one UK-based), North America, Asia, Australia, and Iran. The main anatomical classification system tested was OCSP.(Heuschmann et al. 1999, Asdaghi et al. 2011, Selvarajah et al. 2009, Hand et al. 2006a, Lindley et al. 1993, Dewey et al. 2001) The main mechanistic classification systems tested were TOAST,(Ay et al. 2005a, Cotter et al. 2012, Kolominsky-Rabas et al. 2001, Selvarajah et al. 2009, Yip et al. 1997, Goldstein et al. 2001, Gordon et al. 1993, Meschia et al. 2006, Han et al. 2007, Zhou, Li and Wang 2005, Nam et al. 2012, Wolf et al. 2012) modified versions of TOAST,(Han et al. 2007, Ghandehari et al. 2005, Atiya et al. 2003, Hajat et al. 2001, Fure, Wyller and Thommessen 2005, Johnson et al. 1995) CCS,(Ay et al. 2005a, Ay et al. 2007, Arsava et al. 2010, Marnane et al. 2010, Marnane, Duggan and Sheehan 2009) and A-S-C-O/ASCO1.(Marnane et al. 2010, Cotter et al. 2012, Wolf et al. 2012, Chen et al. 2013) Sample sizes ranged from 14 to 416 (median 54, interquartile range 90). Only six studies tested two or more classification systems in the same population and with the same observers.(Ay et al. 2005a, Cotter et al. 2012, Selvarajah et al. 2009, Han et al. 2007, Nam et al. 2012, Ghandehari et al. 2005) There were no direct within study comparisons of observer reliability of CCS (final computerised version) versus that of TOAST (Table 4.1).

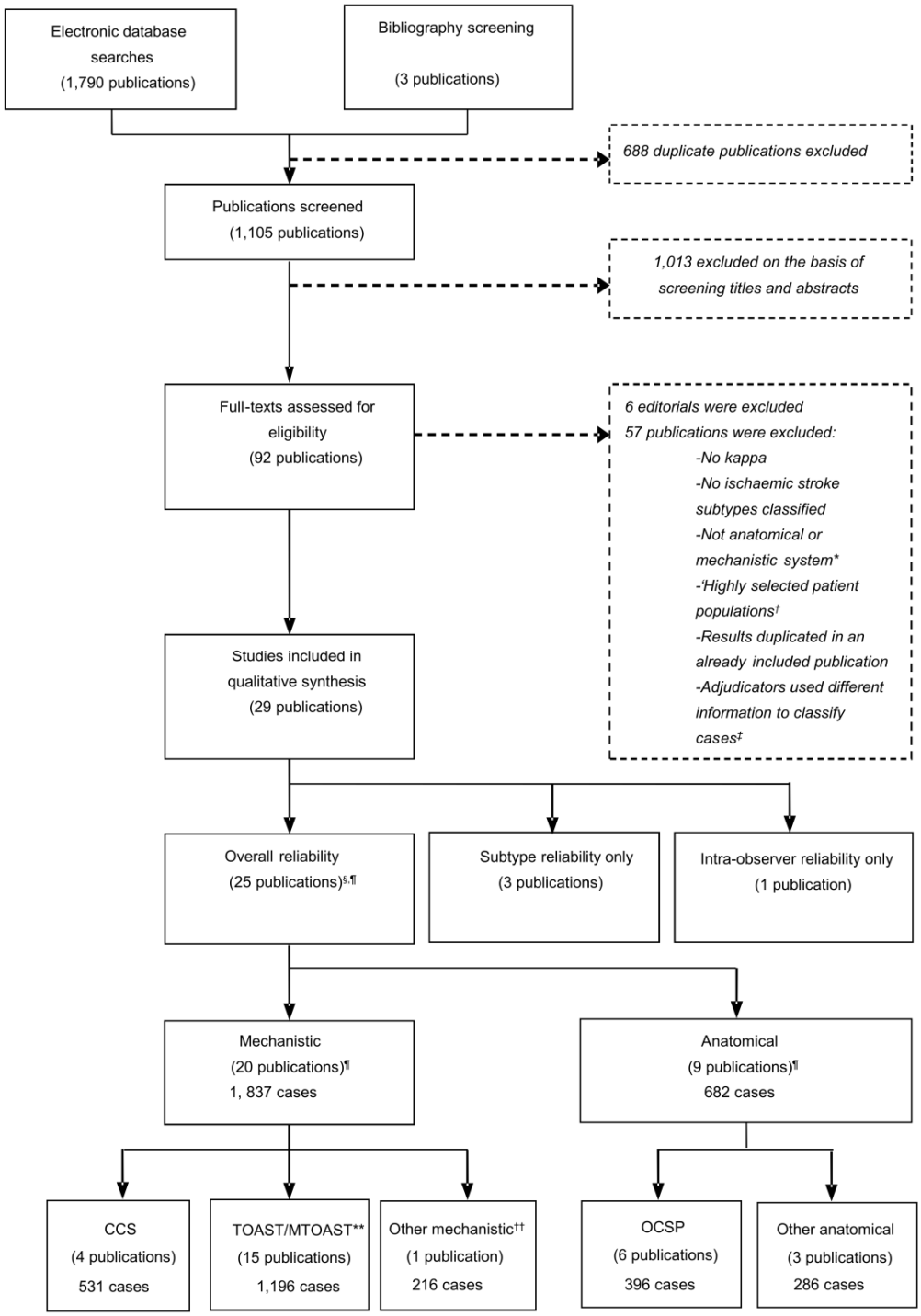

Figure 4.2 Selection of included studies

*Systems which classified stroke severity, disability, or prognosis only, were excluded.

†Two studies classified anterior circulation strokes, one classified strokes occurring post coronary artery bypass graft surgery.

[‡]One study tested classification based on clinical examination versus classification based on a list of symptoms.

[§]Nine of these publications reported 'subtype agreement' in addition to 'overall agreement'.

[¶]Three studies tested more than one mechanistic ischaemic stroke classification system, three tested both an anatomical and a mechanistic classification system.

[**]Eleven studies tested original TOAST, five studies tested modified versions of TOAST (TOAST rules were modified).

[††]Clinicians classified strokes into single cause mechanistic categories without using any rules.

Table 4.1Characteristics of included publications.

Study	Classification system(s) studied	Bias assessment		Study population						Adjudicators			Data presented ‡‡	Time between observations
		Study independent*	Blinding†	Country	Population sampled	Subjects (n)	Mean age	Sample method‡	Time since stroke§	Training¶	Institutions** (n)	Single or consensus††		
Ay 2005	TOAST, CCS§§	×	✓	US	Hospital admissions	50	64	R	-	-	Single	Single	A	-
Ay 2007¶¶	CCS	×	✓	US	Hospital admissions	50	64	R	-	-	Multiple	-	A	5 months
Arsava¶¶	CCS	×	✓	US	Hospital admissions	50	64	R	-	✓	Multiple	Both	A	-
Heusch***	OCSP	✓	-	Germany	Population register	20	73	R	5 days	-	Single	Single	E	4 hours
Asdaghi	OCSP	✓	✓	Canada	Hospital admissions	130†††	-	P	5-10 hours	-	Single	Single	E‡‡‡	-
Marnane 2010§§§	CCS, ASCO1	✓	×	Dublin	Population register	38 CCS 100 ASCO1	70	R	> 6 months	-	Single	Single	A	-
Marnane 2009§§§	CCS	✓	-	Dublin	Population register	38	70	R	-	-	-	-	A	-
Cotter¶¶¶	TOAST, ASCO1	✓	-	UK	Hospital admissions + outpatient clinics	106	49	P	-	-	Single	-	-	-
Kolomin**	TOAST	✓	✓	Germany	Population register	20	73	R	< 7 days to > 7 days	✓	Single	-	E + MR	4 hours

Selvarajah	TOAST, OCSP	✓	✓	UK	Outpatient clinics	90	64	R	14 days	✓	Single	Single	A	-
Yip	TOAST	✓	-	Taiwan	Hospital register	35	-	R	-	-	Single	Single	MR	-
Goldstein	TOAST	×	✓	US	Hospital admissions	14-20	-	R	-	-	Single	Single	MR or A	-
Gordon	TOAST	×	-	US	Hospital register	18	-	R	-	✓	Multiple	Single	A	-
Meschia	TOAST	✓	✓	US	Hospital admissions	30	70	R	-	✓	Multiple	Single	MR	-
Han	TOAST, MTOAST	×	✓	Korea	Hospital register	200	63	R	-	✓	Single	Single	MR	-
Zhou	TOAST	✓	✓	China	Hospital register	300	-	R	-	-	Single	Single	MR	-
Nam	TOAST	×	-	Korea	Hospital register	70	67	R	-	✓	Single	Single	A	-
Wolf	TOAST, A-S-C-O	×	✓	Germany	Hospital admissions	103	69	P	-	✓	Single	Single	MR	-
Ghandehari	MTOAST, Other	-	✓	Iran	Hospital admissions	20	66	R	-	-	Single	Single	A	-
Atiya	MTOAST, Other	✓	✓	US	Self-reported stroke	104-133[****,†††]	>45	R	-	-	Multiple	Single	MR	-
Hajat	MTOAST	×	✓	UK	Multicentre, multi-ethnic register	45	74	R	-	-	Single	Single	A	8 Weeks
Fure	MTOAST	✓	✓	Norway	Hospital admissions	38	66	P	-	✓	Single	Single	E	Same day
Chen	A-S-C-O	✓	✓	China	Hospital register	419	65	P	-	✓	Single	Single	MR	-

Hand	OCSP	✓	✓	UK	Hospital admissions	98[†††]	79	P	<12 hours to >48 hours	×	Single	Single	E	56 minutes
Lindley	OCSP	×	✓	UK	Hospital admissions	85[†††]	-	P	1 day	×	Single	Single	E	Same day
Dewey	OCSP	✓	✓	Australia	Multicentre register	54	-	P	<10 days to >3 months	✓	Single	Single	E	27 hours
Berger	Other	✓	✓	US	Self- reported stroke	216[†††]	40-84	R	-	-	Multiple	Both	MR	6-12 years
Johnson	MTOAST	×	×	US	Multicentre, 'young stroke' register	160	14-44	R	-	-	Multiple	Consensus	A	-
Flossman	Other	✓	✓	UK	Outpatients with MRI	133[†††]	73	P	-	-	Single	Single	E or A	-

MTOAST: TOAST rules were modified. This excludes studies which added data abstraction protocols or computerised algorithms.

[*]Study authors or adjudicators independent of classification system development.

[†]Blinding between adjudicators

[‡] R: Cases selected retrospectively using medical records, P: Cases selected prospectively, meaning that cases were classified shortly after or during hospital admission, and it may have been possible to influence the investigations performed.

[§]Time from stroke to classification. Range, unless otherwise specified.

[¶]Adjudicators were trained prior to the study, or already had experience using the classification system.

[**]Adjudicators from the same (single) or multiple institutions (multiple).

[††]Single: rating is performed by a single adjudicator. Consensus: rating is a consensus between 2 or more adjudicators.

[‡‡]Data presented to adjudicators. MR: adjudicators obtained information from original medical records. E: direct clinical assessment by adjudicator including medical history and examination (with additional access to medical records or test results if required). A: case summaries or data extract prepared by an independent researcher.

[§§]An earlier version of CCS, before the development of a computerised algorithm.

[¶¶]These publications used the same population of cases but different adjudicators in different study settings.

[***]These publications report the same study.

[†††]In these studies a small percentage of cases (<20%) were due to transient ischaemic attack, intracerebral haemorrhage, or stroke mimic.

[‡‡‡]Plus CT brain

[§§§]These two publications report different results from the same study population: overall reliability (2009), subtype reliability (2010).

[¶¶¶] Intra-observer reliability only.

[****]104 cases classified by modified TOAST, 133 classified by 'other' anatomical.

4.3.2 Overall inter-observer reliability

Kappa for overall inter-observer reliability ranged from 0.53 to 0.93 for OCSP, 0.42 to 0.95 for TOAST, 0.49 to 0.91 for modified TOAST systems, and 0.70 to 0.90 for CCS, with no published data for ASCO1 on this reliability measure (Figure 4.3). OCSP and TOAST were assessed in six countries worldwide, including the UK. CCS was assessed in four studies, three US based.(Ay et al. 2005a, Ay et al. 2007, Arsava et al. 2010)

4.3.3 Assessment of bias

Authors or adjudicators were involved in development of the classification system in 11 studies.(Ay et al. 2005a, Ay et al. 2007, Arsava et al. 2010, Goldstein et al. 2001, Gordon et al. 1993, Han et al. 2007, Nam et al. 2012, Wolf et al. 2012, Hajat et al. 2001, Lindley et al. 1993, Johnson et al. 1995) Blinding between observers was not reported by six studies.(Marnane, Duggan and Sheehan 2009, Heuschmann et al. 1999, Cotter et al. 2012, Yip et al. 1997, Gordon et al. 1993, Nam et al. 2012) One or more observers were not blind to the other's classification in two studies, but these did not contribute results for overall inter-observer reliability.(Marnane et al. 2010, Johnson et al. 1995).

Study	Country	Cases (n)	Categories (n)	Ratings* (n)	Study independent†	Blinding‡	Kappa§ (95% CI)
OCSP							
Asdaghi	Canada	130	4	2	✓	✓	0.93¶
Heuschmann	Germany	20	4	2	✓	-	0.56
Selvarajah	UK	90	5	2	✓	✓	0.70
Dewey	Australia	54	4	2	✓	✓	0.53 (0.35-0.70)
Hand	UK	98	4	2	✓	✓	0.58 (0.43-0.72)
Lindley	UK	70	4	2	x	✓	0.59 (0.43-0.74)
TOAST							
Yip	Taiwan	35	5	2	✓	-	0.73
Gordon	US	18	5	24	x	-	0.54
Kolominsky	Germany	20	5	3	✓	✓	0.65 (0.35-0.96)
Meschia	US	30	5	5	✓	✓	0.53
Selvarajah	UK	90	5	2	✓	✓	0.49
Han	Korea	200	5	3	x	✓	0.78
Zhou	China	300	5	2	✓	✓	0.81
Ay 2005	US	50	7	2	x	✓	0.78 (0.64-0.92)
Nam	Korea	70	7	2	x	-	0.69 (0.6-0.78)
Goldstein**	US	14	5	4	x	✓	0.42 (0.32-0.53)
"	"	17	5	2	x	✓	0.54 (0.26-0.82)
"	"	20	5	2	x	✓	0.68 (0.44-0.91)
Wolf	Germany	103	5	2	x	✓	0.95
mTOAST							
Atiya	US	104	5	2	✓	✓	0.49
Han	Korea	200	5	3	x	✓	0.82
Hajat	UK	45	8	5	x	✓	0.91
Fure	Norway	38	4	2	✓	✓	0.88††
Ghandehari	Iran	20	6	3	-	✓	0.88
CCS							
Ay 2005‡‡	US	50	7	2	x	✓	0.90 (0.8-1.00)
Ay 2007	US	50	5	4	x	✓	0.86 (0.81-0.91)
Arsava	US	50	5	11	x	✓	0.80 (0.78-0.81)
Marnane 2009	Dublin	38	5	-	✓	-	0.70

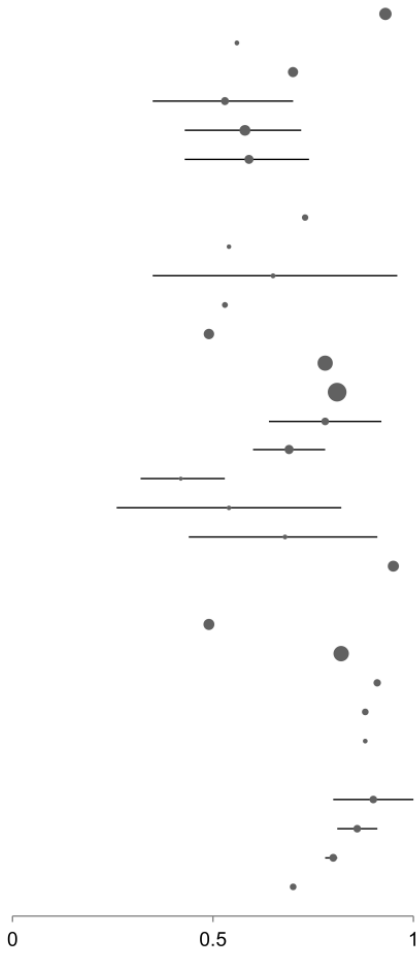

Figure 4.3 Overall inter-observer reliability of the main ischaemic stroke classification systems.*

Figure 4.3 Overall inter-observer reliability of the main ischaemic stroke classification systems. [*]

Results displayed are for the twenty-three studies which assessed overall inter-observer reliability of one of the main ischaemic stroke classification systems (OCSP, TOAST, CCS, Table 4.1).

The area of each circle is proportional to the study size.

Horizontal lines represent 95% confidence intervals (where available).

There were no overall reliability results for ASCO1.

[*]Adjudications were performed by different combinations of observers in different studies. Some are single observations and others are a consensus opinion.

[†]Study was independent if authors and adjudicators were not involved in classification system development.

[‡]Presence or absence of inter-observer blinding.

[§]The displayed value is based on the largest number of cases tested in a single study. If the same cases were tested more than once, an average is displayed, or the simplest classification was chosen (e.g., 5 versus 16 category CCS). Subgroup analyses (based on smaller numbers of cases) are not displayed in this table.

[¶]All patients had CT brain scan and this was used in addition to the clinical syndrome for OCSP classification (31% had signs of early infarction on CT).

[**]14 cases were classified using original medical records, 17 using a data abstraction protocol and a computerised algorithm, and a further 20 using the same computerised algorithm and an updated data abstraction protocol.

[††]Classification was based on clinical examination and results of investigations available on the first weekday after admission to hospital.

[‡‡]An early version of CCS not using computer based assignment.

4.3.4 Factors affecting overall inter-observer reliability

Bias. Based on visual inspection of kappa results (Figure 4.3), there were no convincing effects on reliability of either observer independence from classification system development or blinding, but there were insufficient data to draw reliable conclusions.

Presentation of information. Studies were too heterogeneous to assess the impact on reliability of classification based on direct assessment of the patient, their medical records, or abstracted data.

Number of categories. As expected, within study comparisons found that reliability was consistently higher with fewer categories (Table 4.2).

Clear rules, data abstraction protocols and computerised algorithms. Most classification systems have rules for assigning categories. Only one study included in this review did not have classification rules.(Berger, Kase and Buring 1996) In this study, clinicians used their best judgement to assign cases. Reliability was lower in this study (kappa 0.34) compared to any other study included in this review (kappa range 0.42 to 1.00). Several within study comparisons found better reliability with less ambiguous (CCS versus TOAST) rules,(Ay et al. 2005a) or with a computerised algorithm to classify cases (Table 4.3), but sample sizes were small and 95% confidence intervals overlapped.(Goldstein et al. 2001, Nam et al. 2012) In three US based studies of CCS during different stages of its development, reliability was consistently high (kappa > 0.80), irrespective of data extraction protocols or computer based assignment.(Ay et al. 2005a, Ay et al. 2007, Arsava et al. 2010) There were no within study comparisons of the reliability of the fully developed computerised version of CCS versus TOAST.

Table 4.2 Influence of number of categories on reliability.

Study	System	Cases (n)	Categories (n)	Kappa (95% CI)
Goldstein	TOAST	14	5	0.42 (0.32 to 0.53)
			11	0.29 (0.21 to 0.37)
Ay 2005	CCS*	50	7	0.90 (0.80 to 1.00)
			15	0.86 (0.76 to 0.96)
Ay 2007	CCS	50	5	0.86 (0.81 to 0.91)
			8	0.85 (0.80 to 0.89)
			16	0.80 (0.76 to 0.83)
Arsava	CCS	50	5	0.80 (0.78 to 0.81)
			8	0.79 (0.77 to 0.80)
			16	0.70 (0.69 to 0.71)
	CCS†	50	5	0.81 (0.79 to 0.82)
			8	0.79 (0.77 to 0.80)
			16	0.79 (0.78 to 0.80)
Lindley	OCSP	85	4	0.54 (0.39 to 0.68)
			5	0.43 (0.30 to 0.47)

*An early version of CCS without the computerised algorithm

†Refinements were made to the computer software to include better explanations of terms.

Table 4.3 Influence of data extraction method and computerised algorithms on reliability.

Study		Cases	Classification system	Data extraction method[*]	Kappa[†] (95% CI)
Goldstein[‡]	a)	14	TOAST[§]	Medical record	0.42 (0.32 to 0.53)
	b)	17		Protocol + Computer	0.54 (0.26 to 0.82)
	c)	20		Protocol + Computer	0.68 (0.44 to 0.91)
Nam[¶]	a)	70	TOAST[§]	Abstracted data	0.69 (0.60 to 0.78)
	b)	70		Abstracted data + Computer	0.79 (0.71 to 0.87)
Ay 2005[**]		50	CCS	Abstracted data	0.90 (0.80 to 1.00)
Ay 2007[**,††]		50	CCS	Protocol + Computer	0.86 (0.81 to 0.91)
Arsava[††]	a)	50	CCS	Protocol + Computer	0.80 (0.78 to 0.81)
	b)	50		Protocol + Computer[‡‡]	0.81 (0.79 to 0.82)

[*]Protocol: Data abstracted by an independent researcher using a protocol developed to minimise error. Computer: Classification by computerised algorithm.

[†]If there was a choice based on different numbers of categories, the value corresponding to the smallest number of categories was chosen.

[‡] a), b) and c) are three different sets of cases, classified by different numbers of adjudicators.

[§]If TOAST and a modified version of TOAST were tested in the same study, results are only presented for TOAST, unless the modification was the method of data extraction or the development of a computerised algorithm.

[¶] a) and b) are the same cases classified by the same adjudicators.

[**]These studies were performed within the same research group, but on different case populations.

[††]These studies are based on the same 50 cases but different groups of adjudicators were used.

[‡‡] a) and b) are using the same cases and the same adjudicators. The computer software was updated in b) to reduce ambiguity.

Adjudicator expertise. Three of four within study comparisons (total 197 cases) suggested that adjudicators with a similar level of expertise were more likely to agree than those with different training or expertise, while a fourth, smaller study (20 cases) showed little difference (Table 4.4).(Hand et al. 2006a, Dewey et al. 2001)

Table 4.4 Influence of adjudicator expertise on reliability

Study		Cases	Classification system	Adjudicator expertise*	Kappa† (95% CI)
Hand	a)	98	OCSP	Mixed expertise	0.58 (0.43 to 0.72)
	b)	53‡		Mixed expertise	0.38
	c)	45‡		Less expert only	0.70
Dewey§	a)	54	OCSP	Expert only	0.53 (0.35 to 0.70)
	b)	54		Mixed expertise	0.31 (0.12 to 0.50)
	c)	54		Mixed expertise	0.45¶ (0.26 to 0.64)
Selvarajah	a)	45	OCSP	Mixed expertise	0.70
	b)	45		Expert only	1.00 (1.00 to 1.00)
	c)	45		Mixed expertise	0.90 (0.75 to 1.00)
Ghandehari**	a)	20	Modified	Expert only	0.86 (0.66 to 1.00)
	b)	20	TOAST	Mixed expertise	0.90 (0.75 to 1.00)

* Expert: adjudicators are consultant level stroke specialists, study authors have used the term 'expert' adjudicators, or adjudicators are physicians/neurologists at senior registrar or consultant level. Less expert: Physicians 1-4 years post registration, general practitioners, nurses, or medical students with or without stroke training. Mixed expertise: combination of expert and less expert adjudicators.

†If there was a choice based on different numbers of categories, the value corresponding to the smallest number of categories was chosen.

‡b) and c) are subsets of a)

§a), b) and c) are the same set of cases tested with three different pairs of adjudicators.

¶The expert trained the 'less expert' adjudicator in this pairing, meaning that they are more likely to agree.

**a) and b) are the same set of cases classified by different adjudicators

Adjudicator confidence. Three studies addressed the influence of adjudicator confidence, all on the OCSP classification. In one, agreement was higher when both

adjudicators were highly confident (kappa 0.68), compared to when one or both were less confident (kappa 0.45).(Hand et al. 2006a) In two additional studies, adjudicators recorded their 'best guess' answers when uncertain. This did not appear to affect reliability in one study,(Selvarajah et al. 2009) while in the other, different numbers of categories (reliability with or without use of a 5th 'uncertain' category) precluded a reliable comparison.(Johnson et al. 1995)

Timing of assessment. Two OCSP studies assessed the influence of time from symptom onset to classification.(Hand et al. 2006a, Dewey et al. 2001) Clinical examination (+/- medical record review) was performed at the time of classification. The first study showed that individual history and examination findings were more reliable when classification was performed 12-24 hours after the onset of symptoms (kappa range 0.44 to 0.81), compared to <12 hours (kappa range 0.25 to 0.69) or >24 hours (kappa range 0.19 to 0.51). There were insufficient data in this study to compare the reliability of OCSP classification at different times after the onset of symptoms. The second study found that excluding cases classified >10 days after stroke onset increased reliability of OCSP classification. Another study examined the influence of timing on TOAST, comparing classification in the acute phase (data gathered on admission) to an expert consensus determined later during the admission (data gathered after hospital discharge).(Fure et al. 2005) Agreement was poor (kappa 0.30), reflecting the need for more detailed investigations (available later in the course of admission) for reliable mechanistic classification.

Investigations performed. Overall, investigations performed for TOAST, CCS, and ASCO1 classification were inconsistently reported. Where reported, a high proportion of cases had brain or vascular imaging (80-100%), but a more variable proportion had cardiac investigation (41-100%). From visual inspection of data, the proportion of cases with complete work-up (brain, vascular and cardiac imaging) had no clear effect on reliability. However, if ≥50% cases had more detailed brain, vascular, or cardiac investigation, reliability appeared higher overall (kappa 0.70 to 0.95) than if less detailed investigations were used (kappa 0.49 to 0.65) (Table 4.5).

Table 4.5 Influence of investigations performed on proportion undetermined and overall reliability.

Study	Cases (n)	Overall investigation (%)			Detailed investigation[§] (≥ 50%)	Proportion undetermined[¶] (%)	Reliability (kappa)
		Brain Imaging[*]	Vascular Imaging[†]	Cardiac evaluation[‡]			
TOAST							
Wolf	103	90[**]	100	61	Y	30	0.95
Fure[††]	38	100	100	100	Y	18	0.88
Han	200	100[‡‡]	91	59	Y	36	0.78
Ay 2005	50	98	98	86	Y	61	0.78
Kolominsky[§§]	583	100	89	62	N	41	0.65
Gordon	18	100	83	89	N	16	0.54
Meschia	30	80	100	67	-	30	0.53
Selvarajah[¶¶]	90	100	100	44	N	35	0.49
CCS							
Ay 2005	50	98	98	86	Y	24	0.90
Ay 2007	50	90	82	41	Y	-	0.86
Arsava	50	100	82	41	Y	-	0.80
Marnane 2010	381	99[***]	82	75	Y	26	0.70
ASCO1							
Wolf[†††]	103	90[**]	100	61	Y	31	-
Marnane 2010	381	99[***]	82	75	Y	42	-

[*]This is the proportion of patients who had either CT or MRI. If unknown, the largest proportion investigated (by either CT or MRI) is reported. The proportion of cases which had MRI or diffusion weighted MRI (DWI-MRI) was not clearly reported.

[†]This is the largest proportion investigated by any (or a combination) of the following methods: Doppler ultrasound (USS); CT angiography (CTA); MR angiography (MRA); conventional angiography (angio); transcranial doppler (TCD). The proportion of cases which had evaluation of both intra and extra cranial circulations was not clearly reported.

[‡]ECG and Echocardiogram (ECHO). This is the proportion of patients who had an ECG and either transthoracic or transoesophageal ECHO.

[§]Detailed investigation was defined as ≥50% CTA or ≥50% MRA or ≥ 50% DWI-MRI, or ≥ 50% transoesophageal ECHO. Results left blank if reporting was unclear.

[¶]The undetermined category is made up of three subcategories: unclassified due to multiple potential mechanisms, unknown cause despite full investigation, incomplete investigation.

[**]90% DWI-MRI.

[††]Patients were investigated by protocol to improve completeness of investigation.

[‡‡]66% DWI-MRI

[§§]Cases were selected for classification based on the investigations performed (90% of cases had the investigations required for TOAST classification). The reliability study was based on 20 randomly selected cases from this cohort.

[¶¶]A single researcher classified 301 cases in the original outpatient cohort to derive the proportion undetermined. The 90 consecutive cases chosen for the reliability study are assumed to be representative of the original cohort.

[***]50% DWI-MRI.

[†††]Only 4% of cases were completely undetermined with A-S-C-O classification, meaning that every mechanism had been assigned either grade 0 (disease not present), or grade 9 (insufficient workup).

4.3.5 Proportion of cases undetermined following classification

Most OCSP studies classified all cases.(Asdaghi et al. 2011) However, one (published only in abstract form), reported that 10% of cases were unclassifiable,(Heuschmann et al. 1999) while in two others a 'best guess' answer was included.(Lindley et al. 1993) For mechanistic classifications, proportions of cases unassigned to a single subtype were 10-61% for TOAST,(Heuschmann et al. 1999, Cotter et al. 2012, Kolominsky-Rabas et al. 2001, Zhou et al. 2005) 24-26% for CCS,(Ay et al. 2005a, Marnane et al. 2010) and 31-52% for ASCO1.(Marnane et al. 2010, Cotter et al. 2012, Wolf et al. 2012) These studies varied with respect to country, sample size, and the proportion and type of investigations undertaken. In within study comparisons, CCS had a lower undetermined proportion than TOAST (24% versus 61% in one study,(Marnane et al. 2010) 26% versus 40% in another),(Ay et al. 2005a, Marnane et al. 2010) largely due to a lower proportion of cases unclassified due to multiple potential mechanisms (4% with CCS versus 39% with TOAST in one study).(Ay et al. 2005a) ASCO1 performed similarly to TOAST in within-study comparisons.(Marnane et al. 2010, Cotter et al. 2012, Wolf et al. 2010) Completeness of investigation did not appear to affect the proportion undetermined, although incomplete reporting made it difficult to assess the influence of individual investigations (Table 4.5). Even studies that selected patients based on investigations already completed,(Kolominsky-Rabas et al. 2001) or that used protocols for investigation before classifying by TOAST, modified TOAST or ASCO1, had a substantial undetermined proportion (range 18-41%).(Wolf et al. 2012, Ghandehari et al. 2005, Hajat et al. 2001, Fure et al. 2005)

4.3.6 Subtype inter-observer reliability

For OCSP, within-study comparisons showed consistently higher reliability for assignment of LACS and POCS compared to TACS and PACS (Table 4.6; definitions in Figure 4.1). For single cause mechanistic systems, assignments were consistently most reliable for CE and least reliable for SAO. A-S-C-O (kappa range 0.78 to 1.00),(Wolf et al. 2012, Chen et al. 2013) and the phenotypic CCS (kappa

range 0.79 to 0.95) (Ay et al. 2007) had high reliability for all ischaemic stroke subtypes.

4.3.7 Intra-observer reliability

Only four studies reported intra-observer reliability, and none provided within-study comparisons of different systems.(Ay et al. 2007, Cotter et al. 2012, Selvarajah et al. 2009, Hajat et al. 2001) In between-study comparisons, intra-observer reliability appeared higher for CCS (kappa 0.90)(Ay et al. 2007) and ASCO1 (kappa 1.00)(Cotter et al. 2012) than for TOAST (kappa range 0.48 to 0.93),(Selvarajah et al. 2009, Hajat et al. 2001) modified TOAST (kappa range 0.83 to 0.85)(Hajat et al. 2001) or OCSP (kappa range 0.60 to 0.83),(Selvarajah et al. 2009) but differences between studies in adjudicators, case selection and investigations performed may have contributed to variation in reliability.

***Table 4.6 Reliability of classification to anatomical or single cause mechanistic subtypes*[*]**

Study	Country	Cases (n)	Categories (n)	Adjudicators[†] (n)	Subtype[‡]	Kappa[§]	95% CI
OCSP							
Dewey	Australia	54	4	2	**TACS**	0.45	0.08 to 0.80
					PACS	0.51	0.29 to 0.74
					LACS	0.60	0.35 to 0.84
					POCS	0.51	0.24 to 0.78
				2	**TACS**	0.30	-0.08 to 0.68
					PACS	0.14	-0.12 to 0.40
					LACS	0.48	0.22 to 0.74
					POCS	0.37	0.08 to 0.65
				2	**TACS**	0.46	0.12 to 0.80
					PACS	0.33	0.08 to 0.58
					LACS	0.64	0.39 to 0.88
					POCS	0.41	0.08 to 0.75
Hand	UK	98	4	2	**TACS**	0.64	0.43 to 0.85
					PACS	0.43	0.22 to 0.64
					LACS	0.64	0.43 to 0.85
					POCS	0.70	0.45 to 0.85
TOAST							
Zhou	China	300	5	2	**CE**	0.92	-
					LAA	0.81	
					SAO	0.78	
					Oth	0.90	
Gordon	America	18	5	24	**CE**	0.75	-
					LAA	0.69	
					SAO	0.51	
					Oth	0.53	
Meschia	America	30	5	6	**CE**	0.80	-
					LAA	0.80	
					SAO	0.53	
CCS							
Marnane 2010[‖]	Dublin	38	5	2	**CE**	0.87	-
					LAA	1.00	
					SAO	0.48	
					Other	0.66	
ASCO1							
Marnane 2010[‖]	Dublin	100	5	2	**CE**	0.88	-
					LAA	0.79	
					SAO	0.48	
					Oth	0.66	
Modified TOAST							
Atiya	America	104	5	2	**CE**	0.74	-
					LAA	0.59	
Hajat	London	45	8	5	**CE**	0.84 to 0.97	
					LAA	0.89 to 1.00[#]	
					SAO	0.97	
					Oth	0.16	

Study	Country	Cases (n)	Categories (n)	Adjudicators[†] (n)	Subtype[‡]	Kappa[§]	95% CI
Other Mechanistic							
Johnson	America	160	9	2	**CE**	0.59	0.45 to 0.72
					LAA	0.60	0.42 to 0.74
					SAO	0.28	0.13 to 0.44
					Oth	0.37	0.15 to 0.55

[*]Excludes descriptive mechanistic classification systems.

[†]Adjudications were performed by different combinations of observers in different studies. Some were single observations and others were based on consensus opinion.

[‡]Results are presented for the following subtypes: CE: Cardiac embolism, LAA: large artery atherosclerosis, SAO: small artery occlusion, Oth: Other cause. TACS: Total Anterior Circulation Stroke, PACS: Partial Anterior Circulation Stroke, LACS: Lacunar Stroke, POCS: Posterior Circulation Stroke.

[§]This measures reliability in the classification of an individual subtype versus any other subtype. The displayed value is based on the largest number of cases tested in each study.

| | The second adjudicator was not blinded to the first.

[#] The LAA category was divided into intra and extracranial subcategories. The kappa value of 1.00 was for one case of intracranial LAA.

4.4 Discussion

4.4.1 Summary of findings

My systematic review found widely varying overall inter-observer reliability for OCSP and TOAST, reflecting heterogeneity of study settings, details of how the classification systems were applied, and adjudicators. Most of the studies were small (<100 cases classified), and very few tested more than one system in the same population and with the same observers. Inter-observer reliability for CCS was consistently high, but was not assessed as widely as TOAST or OCSP. Furthermore, the cases included in published studies of CCS underwent detailed investigation, which may have improved reliability. A-S-C-O has been much less extensively studied, and there were no overall reliability results for its single cause counterpart, ASCO1. We found evidence that clear rules, data abstraction protocols, computer based assignment, and fewer categories all improved reliability, and that adjudicators with similar (versus different) levels of expertise were more likely to agree.

Since the last published systematic review of the inter-observer reliability of ischaemic stroke classification (1996),(D'Olhaberriague et al. 1996) existing systems have been more widely studied and newer single cause mechanistic systems (e.g. CCS and ASCO1) have emerged.(Ay 2010, Amarenco et al. 2009a, Chen et al. 2012, Saver 2006) These have attempted to address the major challenge of reducing the proportion of cases undetermined due to multiple potential mechanisms. Amongst a restricted group of studies (identified from our reliability focused search) we found that CCS reduced the proportion undetermined compared to TOAST or ASCO1. Since our search was completed (November 2013) one large multicentre study (~16,000 cases from Europe/America) showed similar proportions undetermined by CCS and TOAST.(McArdle et al. 2014) However, different individual cases were assigned to undetermined categories by TOAST and CCS, and there was a wide range in between system reliability across included centres (which may partly reflect inter-observer variability). A direct comparison of the inter-observer reliability of CCS and TOAST is required to confirm which system (if either) is more consistently reliable, particularly across multiple centres.

By contrast with single cause mechanistic classifications, descriptive mechanistic classification systems (eg. A-S-C-O) classify each potential mechanism separately. They make use of all the available information, allow the contribution of more than one mechanism to be recognised, and record mechanisms unlikely to be related to the incident stroke. However, practical use of A-S-C-O for epidemiological research is limited by the very large number of possible descriptive categories (n=625).

Anatomical classification systems are a practical option for large studies, since they classify all or almost all ischaemic strokes and require no or minimal investigations. However, they are not designed to identify potential underlying mechanisms. OCSP can distinguish between potential large vessel (TACS, PACS, POCS) and small vessel disease (LACS) (albeit with imperfect accuracy),(Asdaghi et al. 2011, Mead et al. 2000, Potter et al. 2010) but not between cardiac embolism and large artery atherosclerosis.

4.4.2 Strengths and limitations

The strengths of my study include: its comprehensive and systematic approach; inclusion of all relevant studies of inter- or intra- observer reliability of anatomical and mechanistic ischaemic stroke classification systems; and exploration of a wide range of factors which might influence reliability, generating findings that are highly informative for large epidemiological studies. There are some limitations. First, it was difficult to assess individual study quality due to incomplete reporting, and the subjective nature of established guidelines (Guidelines for Reporting Reliability and Agreement Studies).(Kottner et al. 2011) Second, the potential association between more detailed investigation of cases and high overall reliability may be due to reporting bias. Third, definitive conclusions were limited by study heterogeneity (meaning that meta-analysis was not appropriate), relatively few reliable within study comparisons of more than one classification system, and no direct comparison of inter-/intra-observer reliability of TOAST and CCS. Fourth, there are likely to be other factors, not assessed here, which influence inter-/intra-observer reliability of ischaemic stroke classification systems. I was only able to assess those factors which had been assessed in within study comparisons. Furthermore, due small numbers of cases, my conclusions based on these within study comparisons are subjective. Fifth,

I did not assess systems without published assessments of reliability (e.g. the recently developed Chinese Ischaemic Stroke Subclassification system).(Chen et al. 2012, Gao et al. 2011) Sixth, my search was reliability focused, meaning that a number of studies which compared the 'proportion of cases undetermined' by ischaemic stroke classification systems have not been included. This limits the strength of my findings regarding proportions undetermined, but it was not the main focus of this review. Finally, it is important to note that reliability is only one aspect of classification system performance for large epidemiological studies. I did not consider the validity of ischaemic stroke classification systems, which would require reference to a 'gold standard' diagnosis.

The newer classification systems CCS and ASCO1, appear promising, but further studies are required to assess their reliability in a wider range of settings, and adaptations to these systems are likely to be required for use in large prospective epidemiological studies. Study characteristics other than the classification system used account for much of the variability in performance, and these characteristics (eg. risk factors of population under study, investigations performed, availability of data and abstraction methods) should be more completely reported.

4.5 Conclusions

Based on the findings of this review, I propose that UK Biobank (and other large epidemiological studies) adopt a flexible approach to ischaemic stroke classification, recognising that no single system is fit for every purpose, and that the system of choice depends on the research question. An anatomical system could be used to classify the majority of cases, irrespective of investigations performed. A mechanistic system could be applied after this, to classify as many cases as possible to a single underlying cause, whilst retaining the data sources which inform the classification. Regardless of the classification system(s) chosen, additional features which appear to enhance reliability should be included, eg. data abstraction protocols, computerised algorithmic assignment, and adjudicator training and assessment. This approach would be flexible, practical, scalable, and resilient to future changes in stroke definitions or health service infrastructure.

4.6 Appendix

4.6.1 Search strategy

EMBASE

1. cerebrovascular disease/ or exp brain infarction/ or exp brain ischemia/ or exp carotid artery disease/ or cerebral artery disease/ or cerebrovascular accident/ or exp occlusive cerebrovascular disease/ or exp stroke/

2. stroke unit/ or stroke patient/

3. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or apoplex$ or isch?emi$ attack$ or tia$).tw.

4. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

5. 1 or 2 or 3 or 4

6. classification/ or exp clinical classification/ or exp disease classification/

7. (type$ of stroke or stroke type$ or subtype$ or classification$ or TOAST or BAMFORD or OCSP or SSS-TOAST or CCS or ASCO).tw.

8. (stroke adj3 categor$).tw.

9. 6 or 7 or 8

10. interrater reliability/ or intrarater reliability/ or observer variation/ or reproducibility/ or predictive value/

11. ((observer or interobserver or inter-observer or intra-observer or intraobserver or interrater or rater or inter-rater or intra-rater or intrarater) adj5 (variation$ or variabilit$ or bias or reliability or agreement or comparison$ or error$ or concordance)).tw.

12. 10 or 11

13. 5 and 9 and 12

MEDLINE

1. cerebrovascular disorders/cl or exp basal ganglia cerebrovascular disease/cl or exp brain ischemia/cl or exp carotid artery diseases/cl or exp cerebral small vessel diseases/cl or exp intracranial arterial diseases/cl or exp "intracranial embolism and thrombosis"/cl or stroke/cl or exp brain infarction/cl or stroke, lacunar/cl or vertebral artery dissection/cl or exp intracranial arteriosclerosis/cl or exp cerebrovascular trauma/cl or exp carotid artery injuries/cl or exp cadasil/cl or exp cerebral arterial diseases/cl or exp leukomalacia, periventricular/cl or exp sneddon syndrome/cl or exp vasculitis, central nervous system/cl

2. cerebrovascular disorders/ or exp basal ganglia cerebrovascular disease/ or exp brain ischemia/ or exp carotid artery diseases/ or exp cerebral small vessel diseases/ or exp intracranial arterial diseases/ or exp "intracranial embolism and thrombosis"/ or stroke/ or exp brain infarction/ or stroke, lacunar/ or vertebral artery dissection/ or exp intracranial arteriosclerosis/ or exp cerebrovascular trauma/ or exp carotid artery injuries/ or exp cadasil/ or exp cerebral arterial diseases/ or exp leukomalacia, periventricular/ or exp sneddon syndrome/ or exp vasculitis, central nervous system/

3. (stroke or cerebrovasc$ or brain vasc$ or cerebral vasc$ or cva$ or apoplex$ or isch?emi$ attack$ or tia$).tw.

4. ((brain$ or cerebr$ or cerebell$ or cortical or vertebrobasilar or hemispher$ or intracran$ or intracerebral or infratentorial or supratentorial or MCA or anterior circulation or posterior circulation or basal ganglia) adj5 (isch?emi$ or infarct$ or thrombo$ or emboli$)).tw.

5. 2 or 3 or 4

6. (type$ of stroke or stroke type$ or subtype$ or classification$ or TOAST or BAMFORD or OCSP or SSS-TOAST or CCS or ASCO).tw.

7. (stroke adj3 categor$).tw.

8. 6 or 7

9. 5 and 8

10. 1 or 9

11. Observer Variation/

12. "reproducibility of results"/

13. "Predictive Value of Tests"/

14. ((observer or interobserver or inter-observer or intra-observer or intraobserver) adj5 (variation$ or variabilit$ or bias or reliability or agreement or comparison$ or error$ or concordance)).tw.

15. 11 or 12 or 13 or 14

16. 10 and 15

Section B: Cross referencing analyses

Chapter 5 Choosing Read codes to identify stroke cases in UK Biobank

- Studies of the determinants of stroke should include only 'first in a lifetime' events, in order to avoid reverse causality.
- Detection of 'first in a lifetime' events requires identification of prevalent cases and exclusion of these participants from the study population.
- In this chapter I report the process of selecting Read codes to identify stroke and its main pathological types in UK Biobank.
- There are limited published data on the accuracy of coded primary care data (Read codes) for stroke
- Informed by my systematic review of coding accuracy (Chapter 2), I searched online resources to select Read codes for a) 'true' stroke cases in UK Biobank and b) additional potential prevalent stroke cases.
- I found that none of the existing online resources had a complete list of Read codes for stroke.
- I concluded that no single code list is fit for every purpose. Care must be taken to select the most appropriate code groups for individual studies.

5.1 Introduction

Follow-up in UK Biobank is primarily through cohort wide linkages to routinely collected coded data from hospital admissions (ICD codes), death certificates (ICD codes), and primary care (Read codes). I previously reviewed published studies of the accuracy of electronic healthcare data for stroke and its main pathological types and recommended the best ICD codes to identify fatal and/or hospitalised stroke cases during follow-up in UK Biobank (Chapter 2). However, I found insufficient data on the accuracy of Read codes for stroke, and therefore no clear evidence to suggest which codes should be used for the ascertainment of non fatal, non hospitalised stroke cases in UK Biobank.

Up to 50% or more of stroke cases are not admitted to hospital in the UK.(Schulz and Rothwell 2003) These cases would be missed if UK Biobank did not include linkages

to primary care coded data (Read codes) in its follow-up strategy. Cohort wide linkage to primary care data should therefore improve the sensitivity of UK Biobank's stroke outcomes detection, in particular for incident non fatal, non hospitalised stroke cases not identified through linkages to hospital admissions and death registration datasets, as well as for prevalent cases not ascertained through participant self-report and retrospective hospital admissions data. In this chapter, I describe how I selected primary care Read codes to search in coded linked primary care data for stroke cases amongst UK Biobank participants.

5.1.1 Read codes

Read codes are a coded thesaurus of clinical terms, and are the most widely used system of clinical coding in UK Primary Care. Read codes are entered by General Practitioners (GPs) to record details of each clinical encounter. These may include a patient's symptoms, examination findings, results of investigations, processes of care, for example referrals to hospital (codes beginning 0 to 9), or a variety of clinical diagnoses (codes beginning A to Z). Multiple Read codes can be entered during a single consultation to capture different aspects of the clinical encounter. Table 5.1 displays the main chapters of the Read clinical classification system.(Davé and Petersen 2009)

Read codes are five character alphanumeric codes. At each level the code may be a lower (small) or upper case (capital) letter or a number. There are 58 available characters at each level and so a theoretical maximum of 656,356,768 available codes.(Chisholm 1990) Coding is hierarchical, meaning that as codes increase in length from 2 to 5 characters, they convey more complicated concepts or more specific clinical diagnoses (Table 5.2).

Table 5.1 Chapters of the Read clinical classification system*

Chapter heading
0 Occupations
1 History & symptoms
2 Examination and signs
3 Diagnostic procedures
4 Laboratory procedures
5 Radiology & physics in medicine
6 Preventative procedures
7 Operations, procedures & sites
8 Other therapeutic procedures
9 Administration
A Infectious and parasitic diseases
B Neoplasms
C Endocrine, nutrition, metabolic and immunity disorders
D Diseases of blood and blood forming organs
E Mental disorders
F Nervous system and sense organ diseases
G Circulatory system diseases
H Respiratory system diseases
J Digestive system diseases
K Genitourinary system diseases
L Complications of pregnancy, childbirth and the puerperium
M Skin & subcutaneous tissue diseases
N Musculoskeletal and connective tissue diseases
P Congenital anomalies
Q Perinatal conditions
R Symptoms, signs and ill-defined conditions
S Injury & poisoning
T Causes of injury and poisoning
U External causes of morbidity and mortality
Z Unspecified conditions

*Codes within each chapter (0 to 9 or A to Z) begin with the same number or letter. For example, codes for Circulatory System Diseases begin with the letter 'G'.

Read codes were first used in the NHS in 1985. Read codes have been updated several times since then, firstly to include more detail (Version 2, released in 1990), and secondly with a more complex structure to allow more coding flexibility (Clinical Terms Version 3, CTV3, released in 1994). The Read clinical classification system is cross referenced to all of the widely used standard classifications such as the International Classification of Diseases (originally ICD-9, but more recently ICD-9-CM and ICD-10), the Office of Population Censuses and Surveys classification of surgical operations and procedures (OPCS 4), the physicians' current procedural terminology (CPT-4), the British National Formulary, and the OPCS classification of occupations.(http://systems.hscic.gov.uk/data/uktc/readcodes) Version 2 Read codes include the Read Drug and Appliance Dictionary (DAAD) which codes medications, appliances, specialist foods, and dressings (and therefore captures primary care prescriptions). However this dictionary has not been maintained for version 3, and has not been updated since 1996.(http://systems.hscic.gov.uk/data/uktc/readcodes/drugandappliancedictionary)

Table 5.2: Example of the Read code hierarchy using individual codes from the Circulatory system diseases chapter.

Level*	Read code	Text definition†
1	G….	Circulatory system diseases
2	G6…‡	Cerebrovascular diseases‡
3	G64..‡	Cerebral arterial occlusion‡
4	G640.‡	Cerebral thrombosis‡
5	G6400‡	Cerebral infarction due to thrombosis of cerebral arteries‡

*The diagnostic complexity increases from level 1 to 5

†This definition is provided with each Read code (from the NHS Read code browser)

‡A single code has been chosen to at each level (2 to 5) to illustrate the hierarchy. Too many codes exist at each level to be able to display them all.

5.1.2 Creating medical code lists to identify cases in primary care databases

Creating code lists for research in primary care databases is a challenging and time consuming process. Redundancy in the Read code classification system means that a

single disease can be identified using many different codes. The process of selecting codes to identify a diagnosis, like stroke, is therefore iterative. Some diseases may be defined easily whereas others are more challenging, requiring combinations of codes representing symptoms, diagnoses, medications, and/or results of investigations.(Springate et al. 2014) Furthermore, the groups of codes selected to identify a particular disease may depend on the research question. Researchers have to decide whether to use broad lists of codes to capture every possible case (maximising sensitivity), or to select codes with narrower definitions in the hope of minimising false positives (maximising PPV).

5.1.3 Online resources to identify Read codes

Research outputs based on UK coded primary care databases are rapidly increasing, but the bespoke 'code lists' used in these studies are generally poorly reported.(Springate et al. 2014) It is recognised that steps are required to improve the quality and reproducibility of research using coded primary care data.(Gulliford et al. 2009, Springate et al. 2014, Davé and Petersen 2009) A number of resources have been developed in recent years which provide researchers with access to ready made code lists for specific diseases, including stroke. Examples include the CALIBER online data portal (CArdiovascular disease research using Linked Bespoke studies and Electronic health Records),(Denaxas et al. 2012) and the more recently developed Clinical Codes online Repository.(Springate et al. 2014) In future, standardisation of these code lists should make the process of Read code selection easier, more transparent, and should also improve comparability between studies using coded primary care data.

5.1.4 Aims

There are at least 202 Read codes for potential stroke.(Gulliford et al. 2009) In this chapter, I aimed to create two Read code groups to identify stroke and its main pathological types.

The first group, called 'acute stroke codes', would be used to identify 'true' diagnoses of ischaemic, haemorrhagic or unspecified stroke. I aimed to include only

the most accurate Read codes in this group (ie. those I predicted have high PPV for stroke and its main pathological types, defined according to WHO definitions).

The second group of codes, called ‘history of stroke codes’ would be used to identify additional UK Biobank participants who may have had a stroke in the past. This would enable exclusion of participants from future population based studies of ‘first in a lifetime stroke’. Such studies should include only ‘first in a lifetime’ events, in order to avoid reverse causality (see Chapter 6.1.1). I used my knowledge of ICD coding accuracy and my own clinical judgement to inform this process.

5.2 Methods

When I began this work, UK Biobank had obtained linked coded primary care data for large subsets of participants in Scotland and Wales. These datasets use version 2 Read codes and so I focussed on these codes (rather than Read codes version 3), for creating my code lists.

5.2.1 Read code search strategy

I used the following resources to generate a comprehensive list of Read codes for 'potential stroke'.

1) Read codes for cerebrovascular diseases provided by the Technology Reference Data Update Distribution Service (TRUD). (https://isd.hscic.gov.uk/trud3/user/guest/group/0/home) These Read codes were provided with cross-maps to ICD-10 codes for cerebrovascular diseases (I60 to I69, G45, and G46).

2) Dictionary of Read codes for cerebrovascular diseases provided by the SAIL (Secure Anonymised Information Linkage) databank. SAIL is an anonymous data linkage system which brings together routinely collected data from the population of Wales for research.(Ford et al. 2009)

3) Quality Outcomes Framework (QOF) indicator sets for stroke, Wales.(http://www.wales.nhs.uk/) These are similar to equivalent lists in England and Scotland. QOF rewards GP practices for the provision of quality care, and measures annual performance against a set of key targets. GP practices are required to maintain a register of patients with stroke/TIA so that secondary prevention targets can be monitored. The QOF indicator sets include lists of Read codes for the diagnosis of stroke/TIA.

4) CALIBER online data portal. CALIBER is a research platform of linked electronic health records (EHR) and administrative health data from primary care, secondary care and disease registries. CALIBER data sources include the Clinical Practice Research Datalink (CPRD), Hospital Episode Statistics England (HES), the Myocardial Ischaemia National Audit Project (MINAP) and mortality/social

deprivation data from the Office of National Statistics (ONS). I reviewed lists of codes for stroke and its main pathological types provided by CALIBER, excluding any codes which were unique to version 3.(http://www.caliberresearch.org/)

5.2.2 Read code selection

1) Acute Stroke

I had previously identified the optimum ICD codes to maximise PPV for stroke (Chapter 2). From my broader list of Read codes for potential stroke (gained above), I selected those which most closely matched the ICD 'acute stroke codes' (Table 5.3, below) using a three-step process.

Table 5.3 International Classification of Diseases codes (ICD codes) for stroke

Acute stroke codes	
ICD-9	430, 431, 434, 436
ICD-9-CM	430, 431, 433.x1, 434, 436
ICD-10	I60, I61, I63, I64

Firstly, I identified Read codes which had been matched by the TRUD service (above) to ICD-10 stroke codes I60, I61, I63 and I64. I reviewed text definitions of these Read codes (see Table 5.2 for examples of text definitions) to check that they made clinical sense for acute stroke. Secondly, after familiarising myself with the Read code hierarchy for stroke (displayed in Appendix 5.6.1), I used the NHS Read code browser (made available on the SAIL platform, hosted by Swansea University) to identify any additional Read codes which matched the ICD-9 stroke codes 430, 431, 434, 436, or the ICD-9-CM stroke codes 430, 431, 433.x1, 434.x1, 436. Table 5.4 shows the NHS Read code browser categories I searched for ICD-9 or ICD-9-CM matching Read codes. I excluded Read codes which matched the ICD-9 code 433 for ischaemic stroke ('occlusion/stenosis of pre-cerebral arteries with or without infarction') because I had previously shown that this was a poorly performing code (PPV 6 to 14% for ischaemic stroke, Chapter 2).

Table 5.4: Read code categories which match ICD-9 codes for acute stroke.

Clinical definition (STROKE)	ICD-9 code (ICD-9-CM)	Read code categories searched* (5 byte version 2)
Subarachnoid haemorrhage (SAH)	430	G60.. and subcategories Gy…and subcategories
Intracerebral haemorrhage (ICH)	431	G61.. and subcategories Gy…and subcategories
Ischaemic stroke (IS)	434 (433.x1, 434.x1)	G64.. and subcategories Gy…and subcategories (Above, plus subcategories of G63..†)
Unspecified stroke	436	G66..and subcategories Gy…and subcategories

*Using the NHS Read code browser

†The Read code G63.. maps to ICD-9 code 433, which has low PPV for ischaemic stroke (range 6 to 14%). Read codes were only selected from the G63.. group if they mapped to the clinically modified ICD-9-CM codes 433.x1 and/or 434.x1, which appear to have a much higher PPVs for ischaemic stroke (range 82% to 96%, Chapter 2).

Finally, I cross-checked my ICD-9 and ICD-10 matched Read code list with the Read code lists for stroke/cerebrovascular diseases provided by SAIL, CALIBER, and QOF. I reviewed the text definitions of any codes which I had not already selected (above) and used my clinical judgement to decide whether or not they should also be included for acute stroke.

2) Past history of stroke:

For codes that might identify additional participants with a past history of stroke, I used a similar process. This time I selected Read codes which matched ICD-10 codes

for ‘sequelae of stroke’ and in particular, codes I69.0 for ‘sequelae of SAH’, I69.1 for ‘sequelae of ICH’, and I69.3 for ‘sequelae of cerebral infarction’. The ICD-9 code 438 ‘sequelae of cerebrovascular diseases’, does not include any specific Read code text definitions for sequelae of stroke, and so I did not include ICD-9 matching Read codes. I used the NHS Read code browser to search the Read code chapters 1 ‘History and symptoms’, 6 ‘Other preventative procedures’ and 9 ‘Administration’, to identify any other potential ‘past history of stroke’ codes based on their text description. As for acute stroke, above, I cross checked my final list with those from SAIL, CALIBER, and QOF to identify any additional potentially relevant codes.

5.3 Results

5.3.1 Acute stroke

There were 241 Read codes related to cerebrovascular diseases in the SAIL dictionary, including codes from chapters 0 to 9 and A to Z, but not including codes for medication prescriptions. Figure 5.1 shows results of the search and selection process, and Table 5.5 shows my final list of 57 Read codes for 'acute stroke', categorised by main pathological type. Fifty Read codes were matched to ICD-10 acute stroke codes by the TRUD service. After reviewing each individual match, I removed one Read code (G65z1 for 'intermittent cerebral ischaemia') because the text description was not in keeping with acute stroke. To the 49 remaining codes, I added 8 by searching the NHS Read code browser. I did not find any additional acute stroke Read codes in the CALIBER or QOF Read code lists. The final Read code list for acute stroke included 20% of the SAIL Read code dictionary for 'cerebrovascular diseases', 76% of the CALIBER 'stroke diagnosed' codes, and 86% of the 'stroke indicator set' QOF codes. No single resource identified all of the acute stroke Read codes. Out of 57 Read codes for acute stroke, 86% were included in the SAIL Read code dictionary for cerebrovascular diseases, 88% were included in the CALIBER code list, and 74% were included in the QOF code list.

Figure 5.1 Selection of Read codes for acute stroke

*For Cerebrovascular Diseases.

†Read codes matched by the Technology Reference Data Update Distribution (TRUD) service to ICD-10 codes for 'other cerebrovascular diseases', not stroke.

‡Read codes which matched ICD-10 acute stroke codes I60, I61, I63 or I64.

§Additional Read codes which matched ICD-9 and ICD-9-CM acute stroke codes 430, 431, 433.x1, 434, or 436.

¶From incident stroke, ischemic stroke and haemorrhagic stroke code lists.

[**]Did not match ICD-9 or ICD-10 acute stroke codes (based on review of text definitions).

[††]Codes not found in the NHS Read code browser (searching 5 byte version 2 codes), these were version 3 Read codes (7 characters long) and therefore not available in the Welsh or Scottish primary care datasets.

[‡‡] These codes were already selected as 'acute stroke codes'.

Table 5.5 Read codes for 'acute stroke' cross mapped to ICD codes for ischaemic, haemorrhagic and unspecified stroke.

Stroke type	**ICD-10**	**ICD-9** (ICD-9-CM)	**READ diagnostic codes** (5-byte Version 2 Read codes used in Welsh dataset)
SAH	I60	430	G60.. and its sub-categories, including G600. to G606./G60X./G60z./Gyu60/Gyu61/Gyu6E.
ICH	I61	431	G61.. and its sub-categories, including G610. to G619./G61X./G61X0/G61X1/G61z./Gyu62/Gyu6F
Ischaemic stroke	I63	434 (433.x1)	G63y0[‡]/G63y1[‡] /G6W../ Gyu6G/ G6400/G6410/ G6X../Gyu63/G6760[‡]/Gyu64/G64z.[‡]/G64z0[‡]/ G64z1[*,‡]/ G64z2[‡]/ G64z3[‡]/G64z4[‡]/G64..
Unspecified stroke	I64	436	G66..[‡]/G667./G668./G669.[‡,§]/G660. to G666.[*,§]

[*]These codes were not present in the TRUD mapped list, and were found by searching the NHS Read code browser for matches to ICD-9 or ICD-9-CM acute stroke codes.

[‡]These codes were not present in CALIBER lists for 'stroke diagnosed'.

[§]These codes were not present in the QOF 'stroke indicator set' of Read codes (Wales).

5.3.2 History of stroke

I identified 18 Read codes for potential past history of stroke, and 40 Read codes for potential past history of Cerebrovascular Diseases (CVD) (Table 5.6). I identified seven of the Read codes for past history of stroke and ten of the Read codes for past history of CVD by cross mapping to ICD-10 codes for 'sequelae of stroke/CVD'. I identified the remainder by searching chapters 0 to 9 of the NHS Read code browser. Table 5.6 displays the Read codes for potential past history of CVD/stroke/main pathological types of stroke. The text definitions of Read codes for past history of stroke and its main pathological types are displayed in Appendix 5.6.2.

Table 5.6 Read codes for 'past history of stroke' and 'past history of CVD' cross mapped to ICD codes for 'sequelae of stroke and its main pathological types' and 'sequelae of CVD'.

Stroke type	ICD-10	ICD-9	READ codes (5-byte Version 2 Read codes used in the Welsh dataset)	
			Diagnosis codes	**Processes of care codes***
SAH	I69.0	-	G680.	14AF.
ICH	I69.1	-	G681.	-
Haemorrhagic stroke	I69.0/I69.1	-	G680./G681.	-
Ischaemic stroke	I69.3	-	G683.	-
Unspecified stroke	I69.4	-	G68X./Gyu6C	14A7./14AK./7P242/8HHM./1M4../661M7/662M1/662M2/8IEC./662e./
Sequelae of CVD†	I69	438	G680./G681./G683./ G68X./Gyu6C/ G68../G682./Gyu6B/G68W./Gyu6D./	14AB./14AB0/ZV12D/ 388I./1477./8E24./14A7./ 14AK./662M./662e./7P242/ 8HHM./ZV125/1M4../ 661M7/662M1/662M2/ 8IEC./9Om../13YA./9Om0./ 9Om1./9Om2./9Om3./ 9Om4./6F…/8E24./14AF./ 662o./Q412./

SAH= Subarachnoid haemorrhage, ICH= Intra-cerebral haemorrhage, CVD=Cerebrovascular diseases.

*Including Read codes from chapters 1 for History and Symptoms, 6 for Preventative procedures, 7 for Operations, procedures and sites, 8 for other therapeutic procedures, and 9 for Administration.

†These are codes for 'non stroke CVD' and TIA.

5.4 Discussion

In this chapter I used a variety of different resources to identify more than 200 Read codes for potential cerebrovascular diseases. I then selected 57 Read codes for 'acute stroke', and 18 Read codes for 'past history of stroke', based on cross mapping to ICD codes for 'acute stroke'/'stroke sequelae', and clinical review of Read code text definitions. By identifying codes from a number of different resources, and by selecting only those codes which matched the best ICD codes for acute stroke, I hope to have created a comprehensive Read code list with high PPV for acute stroke. In future, the Read code groups I have selected will be used to identify new cases of stroke during follow-up in UK Biobank, and/or to identify participants in UK Biobank who have had a stroke in the past. These potential cases should be validated by expert led adjudication to determine the PPV of Read codes for stroke and its main pathological types.

An advantage of linkage to Read codes during follow-up in UK Biobank is the potential to identify additional non fatal, non hospitalised stroke cases which would not be detected by ICD codes or participant self-report. This approach (combining multiple different coded data sources) should improve the sensitivity of stroke case ascertainment, and increase the numbers of stroke cases included in future nested case cohort or case control studies of the associations between risk factors and stroke. It should also reduce the risk of outcome selection bias (under selection of milder stroke cases, which are less likely to be admitted to hospital). Whilst it is important to include as many cases as possible during follow-up, false positive cases should be kept to a minimum (misclassification of outcomes reduces power to detect differences between them in the strength of their risk factor associations). In selecting my 'acute stroke codes' I have therefore focussed on selecting individual Read codes to maximise PPV for stroke. If I had included Read codes with high sensitivity for potential stroke, I would risk introducing too many false positive cases, and therefore too many potential cases for subsequent adjudication.

In other circumstances it may be more important to create code groups which maximise the number of potential stroke cases detected, placing less importance on PPV. In creating my 'history of stroke group', I aimed to detect additional

participants who have had a stroke in the past (in addition to those detected using ‘acute stroke codes’ to search for prevalent stroke). Combining ‘history of stroke codes’ with ‘acute stroke codes’, should enable more complete exclusion of participants with a potential history of stroke from future studies of ‘first in a lifetime stroke’. Although this approach might exclude some ‘true’ ‘first in a lifetime cases’ from follow-up, it would minimise inclusion of ‘false positive’ ‘first in a lifetime cases’. The approach could be extended further by searching for and excluding any participant from such a study who had any of the ‘cerebrovascular diseases’ codes.

Although bespoke Read codes lists are being developed for use in research,(Denaxas et al. 2012, Springate et al. 2014) they have not yet been validated for specific diseases, like stroke. Researchers still need to review individual codes for inclusion in any given code list and select the most relevant codes for their research question. In this chapter I showed that only 20% of Read codes in the SAIL dictionary for ‘potential CVD’ matched validated ICD codes for acute stroke. Furthermore, only 76% of CALIBER stroke codes, and only 86% of QOF stroke codes matched validated ICD codes for acute stroke. Many of the codes excluded from these lists were more likely to identify ‘old’ cases of stroke, or cases of ‘other cerebrovascular disease’, than ‘acute stroke’ diagnoses. I also found that no single resource (TRUD service, CALIBER list, or QOF code list) included all of the potentially relevant Read codes for ‘acute stroke’.

The strengths of my approach therefore include searching multiple code lists to identify as many potentially relevant ‘stroke’ codes as possible, and selecting the most appropriate codes from within these lists to maximise PPV for acute stroke. Rather than relying on clinical judgement alone, I selected Read codes for acute stroke which best matched validated ICD codes (using knowledge gained in from a systematic review of the accuracy of routinely available coded data for stroke, Chapter 2). A study published in 2009 demonstrated poor inter-observer agreement in the selection of Read codes for stroke, suggesting that clinical judgement alone may be a poor basis for generating an accurate and reproducible code list.(Gulliford et al. 2009) Four independent clinical researchers were asked to specify which Read codes they felt were likely to represent an ‘acute stroke’ from amongst a list of 202

potential codes. Overall agreement between all four researchers was low (kappa 0.23). Defining the disease outcome, in this case by selecting the most appropriate codes to maximise PPV for stroke, is a crucial part of study design. It is not only important for transparency and reproducibility, but it may also influence the results of studies which use stroke as an exposure or an outcome. For example, different groups of Read codes selected to represent 'acute stroke' have been shown to provide different estimates of one year post stroke mortality.(Gulliford et al. 2009)

5.5 Conclusions

In conclusion, I have selected groups of Read codes which will be used to identify potential incident and prevalent stroke cases in UK Biobank (see Chapter 6). I have focused on identifying groups of Read codes to maximise PPV for stroke, using version 2 Read codes which will be applied to the linked Scottish and Welsh primary care datasets. It will also be necessary to compile similar lists for use in the English primary care datasets, based on Read codes version 3. Different groups of codes are likely to influence the outcomes of research studies based on coded healthcare data. In future, wider groups of Read codes could be selected to maximise sensitivity for non fatal, non hospitalised stroke cases in UKB (analogous to the use of all cerebrovascular ICD codes, I60-I69, G45 and G46 to maximise sensitivity for hospitalised or fatal stroke cases, Chapter 2), but these potential cases would require further adjudication. In the long run it may be possible to conduct sensitivity analyses in UK Biobank to see how associations between parameters, like BP and stroke, vary depending on the groups of codes selected to define stroke (and/or its main pathological types), as well as on the sources of these codes (ie. use of coded primary care data in addition to coded hospital and death certificate data). In the next chapter I present numbers of prevalent and early incident stroke cases in UK Biobank, identified using combinations of coded data sources (Read codes, ICD codes, and/or self-report).

5.6 Appendix

5.6.1 The Read code hierarchy for cerebrovascular diseases

G…. = Circulatory system diseases

G6… = Cerebrovascular disease

G60.. = SAH

G61.. = ICH

G62.. = Other and unspecified intracranial haemorrhage

G63.. = Pre-cerebral arterial occlusion

G64.. = Cerebral arterial occlusion

G65.. = Transient cerebral ischaemia

G66.. = Stroke and CVA unspecified

G67.. = Other cerebrovascular diseases

G68.. = Late effects of cerebrovascular disease

G6W..= Cerebral infarction due to unsp. occlusion/stenosis of precerebral arteries

G6X.. = Cerebral infarction due to unsp. occlusion/stenosis of cerebral arteries

G6y.. = Other specified cerebrovascular disease

G6z.. = Cerebrovascular disease nos.

Gy… = Other specified diseases of the circulatory system

Gyu.. = [X] Additional circulatory disease classification terms

Gyu6. = [X] Cerebrovascular diseases

5.6.2 Text definitions of past history of stroke Read codes

G680. = Sequelae of SAH

14AF. = History of SAH

G681. = Sequelae of ICH

G683. = Sequelae of IS

G68X. = Sequelae of stroke not specified as haemorrhage or infarction

Gyu6C = Sequelae of stroke not specified as haemorrhage or infarction

14A7. = History of CVA/stroke

14AK. = Stroke in the last year

7P242 = Delivery of rehabilitation for stroke

8HHM. = Referral to multidisciplinary stroke function improvement service

1M4.. = Central post stroke pain

661M7 = Stroke self-management plan agreed

662M1 = Stroke 6 month review

662e. = Stroke/CVA annual review

662M2 = Stroke initial post discharge review

8IEC. = Referral to stroke multidisciplinary improvement declined

Text definitions of past history of cerebrovascular diseases Read codes

14AB. = History of TIA

14AB0 = History of amaurosis fugax

ZV12D = Personal history of TIA

1477. = History of cerebrovascular disease

8E24. = Dysphasia training

14AK. = History of stroke in past year

662M. = Stroke monitoring

662e. = Stroke annual review

7P242. = Delivery of rehabilitation for stroke

8HHM. = Referral to multidisciplinary stroke function improvement service

ZV125 = Personal history of cerebrovascular accident

1M4.. = Central post stroke pain

661M7 = Stroke self-management plan agreed

662M1 = Stroke six month review

662M2 = Stroke initial post discharge review

8IEC. = Referral to stroke multidisciplinary function improvement service declined

9Om.. = Stroke/transient ischaemic attack monitoring administration

13YA. = Stroke group member

9Om0. = Stroke/transient ischaemic attack monitoring first letter

9Om1. = Stroke/transient ischaemic attack monitoring second letter

9Om2. = Stroke/transient ischaemic attack monitoring third letter

9Om3. = Stroke/transient ischaemic attack monitoring fourth letter

9Om4. = Stroke/transient ischaemic attack monitoring fifth letter

6F… = Stroke prevention

8E24. = Dysphasia training

662o. = Haemorrhagic stroke monitoring

Q412. = Perinatal subarachnoid haemorrhage

Chapter 6 Identification of prevalent and early incident stroke cases in UK Biobank

- Up to 50% of stroke cases in the UK are not admitted to hospital. These cases would be missed if UK Biobank did not included linkages to coded primary care data.
- In this chapter I report the contribution of multiple overlapping sources of coded healthcare data to the identification of early incident and prevalent stroke cases in UK Biobank
- Analyses were performed in the UK Biobank population (~500,000), and in a sub-cohort with complete linkage to coded primary care data (~22,000)
- I found that the majority of prevalent stroke cases were identified by participant self-report, and were mostly stroke of 'unspecified type'.
- Incident cases detected by coded hospital or death certificate data were mostly specified pathological type (ischaemic or haemorrhagic stroke).
- Coded primary care data identified ~38% of incident strokes in the UKB sub-cohort, half of which were not detected by any other data source.
- I concluded, as expected, that linkage to coded primary care data should improve the completeness of UKB stroke outcomes ascertainment.
- In future, a large proportion of incident stroke cases in UKB may be classified into pathological types using coded data alone.

6.1 Introduction

In this chapter I present the numbers of prevalent and early incident stroke cases which were identified in UK Biobank using combinations of linked coded healthcare data.

6.1.1 Definition of prevalent, incident, and recurrent stroke cases

In order to investigate new potential risk factors for stroke, or to better understand the associations between established risk factors, stroke, its main pathological types, and subtypes, it is necessary to conduct well designed prospective studies where the

exposure of interest is measured before the occurrence of the outcome. In UK Biobank, nested case cohort or case control studies will be used to explore associations between exposures identified at the UK Biobank baseline assessment and first ever stroke events identified during follow-up. Data on a wide variety of potential exposures were collected during the UKB baseline assessment. For each participant, the time prior to the baseline assessment is the exposure period, and the time after the baseline assessment is the follow-up period (Figure 6.1). It is important to identify only those participants who have had their first ever stroke (ie. first event in a lifetime) during follow-up. Participants who have had a previous stroke will need to be excluded to avoid reverse causality (ie. development of the risk factor after the occurrence of disease) which may bias studies of associations between risk factors, like blood pressure, and stroke. Therefore, for the purpose of this work, incident stroke is defined as any first in a lifetime event, occurring after recruitment to UK Biobank, prevalent stroke is defined as any stroke which has occurred prior to UK Biobank recruitment, and recurrent stroke is defined as any subsequent event in an individual who has had a stroke before (irrespective of the timing of this event). Recurrent stroke is excluded from the count of incident and/or prevalent stroke cases. The distinction between prevalent, incident, and recurrent stroke is illustrated in Figure 6.1.

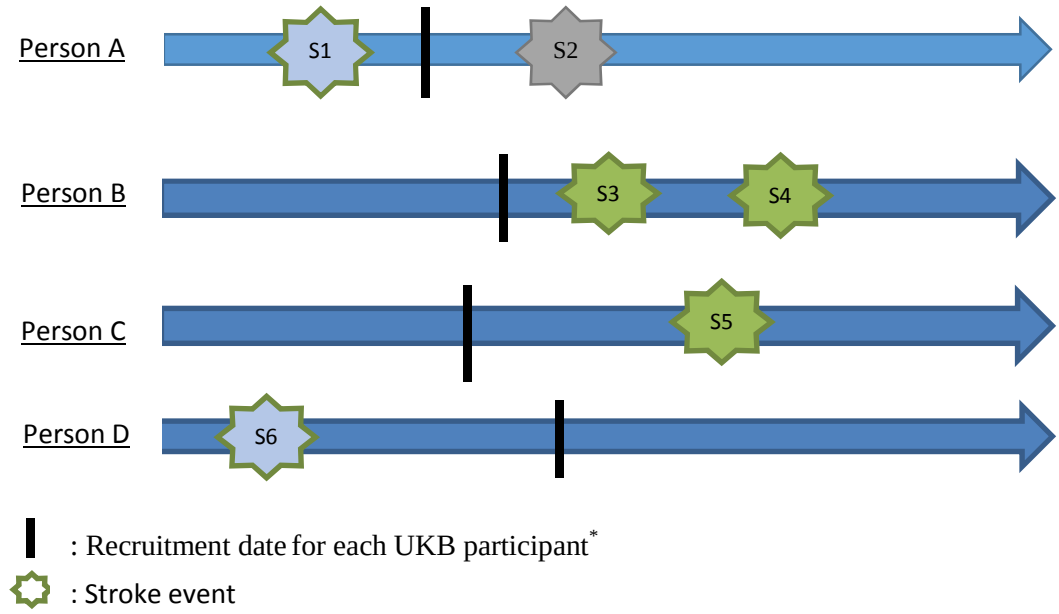

Figure 6.1 Illustration of prevalent, incident, and recurrent stroke in UK Biobank†

S1: Prevalent stroke, S2: recurrent stroke, S3: incident stroke, S4: recurrent stroke, S5: incident stroke, S6: Prevalent stroke.

Person B and Person C would be identified as individuals with incident stroke for inclusion in a nested case cohort or case control study of the associations between risk factors and stroke in UK Biobank. Events S3 and S5 would be used.

*Each individual's baseline assessment date.

†Produced with permission from Q. Zhang and C. Sudlow.

6.1.2 Data linkage in UK Biobank

UK Biobank has planned cohort wide linkages to coded healthcare data from hospital admissions, death certificates, and general practices across England, Scotland, and Wales. At the time of this study, linkages were not complete for all three data sources in all three countries (ie. linked coded data were not available for all participants). Table 6.1 shows the numbers of participants for whom linked coded data were available in UK Biobank across England, Scotland, and Wales, and the periods of data coverage, stratified by the coded data source used.

Table 6.1: UK Biobank data linkages.

Coded data source	Period available*	Data linkages† (% of 502,523)	Participants with records‡ (n)
Baseline assessment	2006 - 2010	100	502,523
Death certificates	2006 - 2014	100	8,662
Hospital admissions	1980 - 2013	100	352,730
Primary care records¶	1940 - 2012	~11	22,562

*Time period for which coded data were available

†Proportion of participants (out of 502, 523 in the whole UK Biobank cohort) for whom data linkage was complete

‡Number of participants (out of those with linked data) who had at least one coded record.

¶Linkages to primary care coded data were complete for only ~40% of UKB Welsh and Scottish participants, which represented ~11% of the UK Biobank cohort.

Data from the UKB baseline assessment were collected between 2006 and 2010 (the date of recruitment varied per participant), coded hospital admission records (ICD-9 or ICD-10 codes) from England (Hospital Episode Statistics, HES), Scotland (Scottish Morbidity Record, SMR01), and Wales (Patient Episode Database for Wales, PEDW), were available for the whole UK Biobank cohort from 1980 to 2013, and national death registry data (ICD-9 or ICD-10 codes based on death certificate diagnoses) were available for the whole cohort until the end of 2014. Coded primary care data were only available for around half of UKB Welsh and Scottish participants (~56,663 participants were recruited in Scotland and Wales) until the end of 2012.

6.2 Aims

The aims of this chapter are to determine the numbers of prevalent and incident stroke events, and the equivalent numbers of each of the main pathological types of stroke (ischaemic stroke, subarachnoid haemorrhage, SAH, and intracerebral haemorrhage, ICH) which are identified by different coded data sources (participant self-report, ICD codes, and Read codes) across England, Scotland and Wales. Linkage between these datasets will be used to determine the numbers of cases

identified by any single data source amongst UK Biobank participants, as well as the numbers of cases identified by combinations of coded data. This novel work should determine the potential contribution of each separate data source to the identification of stroke and its main types in UK Biobank. In particular, it will determine the number of stroke cases which can be classified into a main pathological type (ischaemic or haemorrhagic stroke) based on coded data alone. It will also demonstrate the numbers of non fatal, non hospitalised stroke cases which are identified using coded primary care data, thereby illustrating the important potential of this data source to improve the accuracy and completeness of stroke outcomes identification.

6.3 Methods

6.3.1 Single data sources

We initially investigated each of the four data sources separately. We counted the numbers of incident and prevalent stroke cases identified using coded hospital, death certificate, and primary care data, as well as the number of prevalent stroke cases identified by participant self-report (at the UK Biobank baseline assessment). The denominator population for each of these analyses was the total number of participants who had been linked to each coded data source (Table 6.1). The optimum ICD and Read codes to identify 'acute stroke' and each of its main pathological types have been described in previous chapters (Chapters 2 and 5). Table 6.2 displays the ICD codes for 'acute stroke' and its main pathological types, along with the codes selected from the UK Biobank baseline assessment to identify self-reported stroke. Read codes for 'acute stroke' and its main pathological types are displayed in Chapter 5 (Table 5.5).

Table 6.2 Codes for 'acute stroke' and its main pathological types

Diagnosis sought	**ICD-10**	**ICD-9**	**UKB self-report codes***
Subarachnoid haemorrhage (SAH)	I60	430	1086
Intracerebral haemorrhage (ICH)	I61	431	1491
Ischaemic stroke (IS)	I63	434	1583
Stroke, unspecified	I64	436	1081
Stroke (all types)	I60-I64	430,431,434,436	1081,1086,1491,1583

*Codes for self-reported stroke and its main pathological types extracted from the UK Biobank baseline assessment dataset. Self-reported strokes were included from the touchscreen questionnaire and the nurse-led interview.

The same groups of ICD and Read codes ('acute stroke codes') were used to identify incident and prevalent stroke cases. The definition (incident versus prevalent stroke) depended on the timing of the coded event in each separate data source with respect to UKB recruitment (shown in Figure 6.1). For simplicity, codes for 'history of stroke' (described in Chapter 5), were not used in these analyses. For cases identified

using ICD coded data we also determined the numbers of events which were coded in the primary diagnostic position versus the secondary diagnostic position. (See Chapter 2 for definitions of primary versus secondary diagnostic positions).

6.3.2 Multiple data sources

We then analysed the contribution of multiple overlapping sources of data to the identification of incident and prevalent stroke cases. We performed these analyses in the whole UK Biobank cohort (without coded primary care data), and in a sub-cohort of the UKB population (with coded primary care data).

6.3.3 Whole UK Biobank cohort.

We used combinations of coded hospital, death certificate, and/or self-report data to identify early incident and/or prevalent stroke cases in the whole UK Biobank cohort. Linkage was complete for all three data sources, but the periods of data coverage (the periods of time for which coded data were available) differed from one data source to another (Table 6.1). We truncated the search for 'prevalent stroke' (using linkages to coded hospital data), to the date of UK Biobank recruitment (which varied by individual). Self-report data were ascertained at the date of recruitment, and could only be used to identify 'prevalent' cases. We truncated the follow-up period for 'incident stroke' (using linkages to either coded hospital or death certificate data), to the end of 2013. Primary care data were not included in these analyses because linkage was only complete for ~11% of the UKB cohort (Table 6.1).

6.3.4 UK Biobank sub-cohort

We created a sub-cohort of ~ 22,500 participants for whom linked coded data were available for all four data sources (Table 6.1). In this cohort we used combinations of coded hospital, primary care, and self-report data to search for 'prevalent' stroke (truncated to the date of UKB recruitment, as described in 6.4.1). We used combinations of coded hospital, death certificate, and primary care data to search for incident stroke. For comparability, follow-up of incident stroke was truncated to the end of December 2012 (Table 6.1).

6.3.5 Identifying the main pathological type of stroke

We determined the numbers of ischaemic, haemorrhagic (ICH and/or SAH), and unspecified strokes identified by each separate data source. An incident stroke of confirmed pathological type (ischaemic stroke, ICH, or SAH) was the first ever event of that type in a lifetime. A participant could have an incident haemorrhagic stroke (SAH or ICH), despite having a past history of ischaemic stroke, and an incident ischaemic stroke, despite having a past history of haemorrhagic stroke.
Furthermore, it was theoretically possible to have multiple different stroke codes for the same hospital admission, either because the main pathological type of stroke, ischaemic versus haemorrhagic stroke, was determined later in the course of a hospital admission (therefore an unspecified stroke code was later followed by a specified stroke code), or because the participant had an early stroke recurrence during their hospital stay (ie. an ischaemic, haemorrhagic or unspecified stroke was followed by a recurrent stroke of any type). It was not possible to distinguish between these alternatives based on the coded diagnoses alone.

We created the following rules to deal with multiple different coded diagnoses within the same hospital admission. Firstly, a single period of inpatient stay is divided into multiple hospital episodes. The episode is a period of inpatient hospital stay under the care of a specified hospital consultant, or within a particular hospital department. Each episode includes one primary and up to 19 secondary ICD codes, which represent the main condition(s) responsible for that period of inpatient stay. Codes within a single episode are given the same event dates. Therefore, when multiple stroke codes were present in sequential episodes, we selected the earliest stroke code (the code associated with the earliest episode within that admission) to represent incident stroke, and the earliest stroke code of specified pathological type to represent incident ischaemic or haemorrhagic stroke. Secondly, when more than one stroke code was entered for a single hospital episode, therefore with same event date (ie. one type of stroke code in the primary diagnostic position and another type of stroke code in the secondary diagnostic position), we selected the primary position stroke code to represent the diagnosis for that admission.

6.3.6 Characteristics of selected participants

We explored baseline characteristics of participants in whole UK Biobank cohort, compared to the population recruited in Scotland and Wales (n=56,663, of which ~40% made up our sub-cohort). Characteristics were extracted from the linked UK Biobank baseline assessment dataset including age at baseline (≤44, 45-49, 50-54, 55-59, 60-64, 65+), gender, smoking status, Townsend deprivation index and comorbidities (including Stroke, TIA, other CVD, heart disease, diabetes, and renal disease).

Statistical analyses were conducted by a statistical epidemiologist, Dr Qiuli Zhang, in SAS.9.2.

6.4 Results

6.4.1 Prevalent stroke

Amongst 502, 523 participants in the whole UK Biobank cohort, 8,769 (~1.7%) had at least one stroke prior to UKB recruitment based on cohort wide linkages to participant self-report and coded hospital data (Figure 6.2). The majority of prevalent stroke cases were identified by participant self-report (~95%). Coded hospital data (ICD codes) identified an additional 465 participants with prevalent stroke (the remaining 5% of prevalent stroke cases), but only 2,786 participants with prevalent stroke overall, and only ~26% of participants who had self-reported stroke at the UKB baseline assessment. The prevalence of stroke in UK Biobank was ~1.6% based on participant self-report, and ~ 0.6% based on coded hospital data (ICD codes).

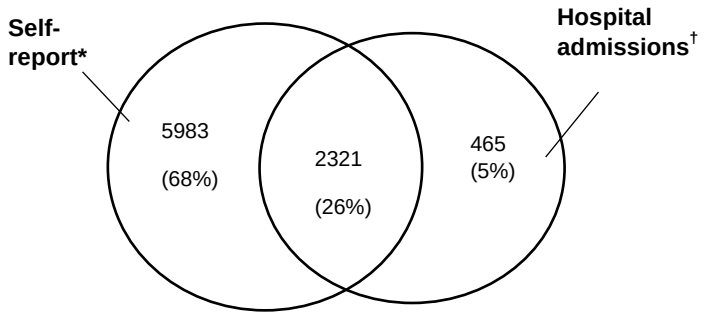

Figure 6.2 Participants with prevalent stroke (n=8,769) identified by self-report and coded hospital admissions data (ICD codes) amongst 502, 523 in the whole UK Biobank cohort.

*Stroke identified by participant self-report at the UK Biobank baseline assessment

†ICD codes I60, I61, I62, I63, I64, 430, 431, 434, 436

Amongst 22,562 participants in the UKB sub-cohort, 1,145 (~5%) had at least one stroke prior to UKB recruitment based on cohort wide linkages to coded primary care data (Read codes), coded hospital data (ICD codes) and self-report (Figure 6.3). Again, the majority of these cases were identified by participant self-report (~90%). An additional ~10% of prevalent stroke cases were detected using a combination of coded primary care data (Read codes) and coded hospital data (ICD codes). The prevalence of stroke in this sub-cohort was ~4.6% based on self-report, ~1.8% based on coded hospital admissions data, and ~1% based on coded primary care data.

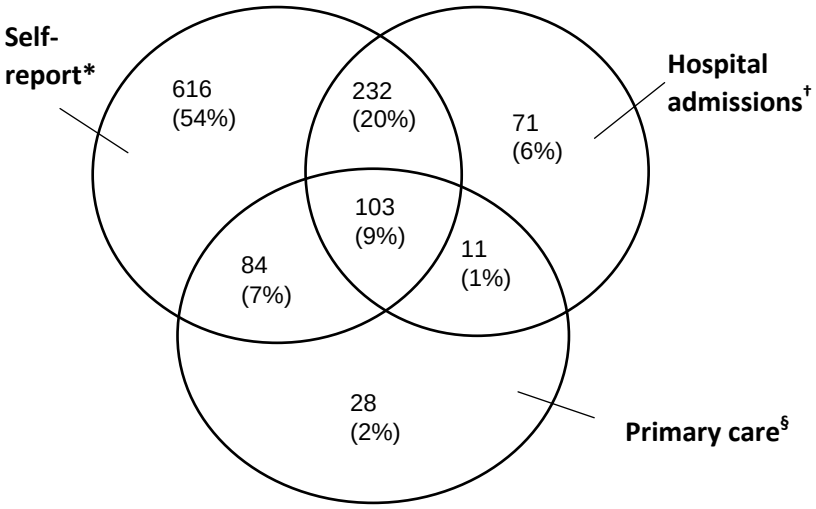

Figure 6.3 Participants with prevalent stroke (n=1,145) based on self-report, coded hospital admissions data (ICD codes), and coded primary care data (Read codes) amongst 22,562 in the UKB sub-cohort with complete linked data.

*Stroke identified by participant self-report at the UK Biobank baseline assessment

†ICD codes I60, I61, I62, I63, I64, 430, 431, 434, 436

§Read codes displayed in Chapter 5 (Table 5.5 'acute stroke' codes)

6.4.2 Incident stroke

There were 850 incident stroke cases in the whole UK Biobank cohort using linkages to ICD coded data from hospital admissions and/or ICD coded data from death certificates until the end of December 2012 (Figure 6.4). The majority of these (~92%) were detected using hospital ICD codes. An additional 8% were detected using death certificate ICD codes. Only 3% of incident cases were detected using both data sources.

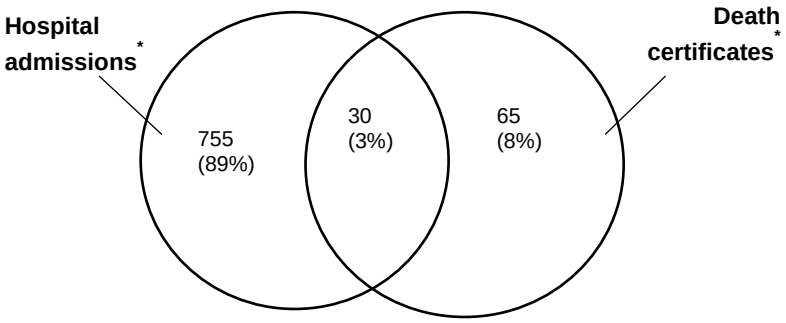

Figure 6.4 Incident stroke cases (n=850) identified by coded hospital data (ICD codes) and coded death certificate data (ICD codes) amongst 502, 523 UK Biobank participants

* ICD codes I60, I61, I62, I63, I64, 430, 431, 434, 436

Amongst 22,562 participants in the UKB sub-cohort, 186 incident stroke cases were detected using linkages to ICD coded data from hospital admissions and/or death certificates, and/or linkages to primary care coded data (Read codes) until the end of December 2012 (Figure 6.5). Around 19% of all cases detected were present in primary care coded data only. Of all cases detected (n=186), ~11% were identified using death certificate ICD codes, ~78% were identified using hospital ICD codes, and ~38% were identified using primary care Read codes. No cases were identified by all three data sources simultaneously. The proportion of cases identified from hospital admissions which were also identified by death certificates, or also identified in the primary care record, were ~10% and ~24%, respectively.

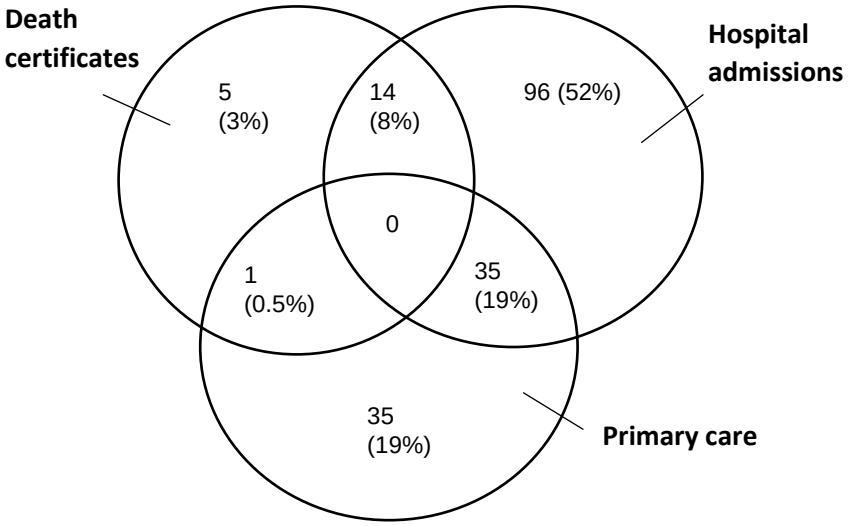

Figure 6.5 Incident stroke cases (n=186) identified by coded hospital data (ICD codes), coded death certificate data (ICD codes), and coded primary care data (Read codes) amongst 56,663 in the UKB sub-cohort with complete linked data.

6.4.3 Coding patterns

The majority of stroke cases identified using coded hospital data were recorded in the primary diagnostic position (97%). Similarly, the majority of fatal strokes identified using coded death certificate data were recorded as the underlying cause of death (81%). Proportions were similar irrespective of region (England, Scotland, or Wales) and irrespective of the pathological type of stroke (ischaemic stroke, SAH, or ICH). Figure 6.6 shows the proportions of ICD-10 codes which were in the primary diagnostic position amongst all coded hospital admissions for stroke in England (HES), Scotland (SMR), and Wales (PEDW). Results are also displayed for each of the main pathological types of stroke (Subarachnoid haemorrhage, SAH, Intracerebral haemorrhage, ICH, and Ischaemic stroke, IS), as well as stroke, unspecified. In the coded hospital data, only 1.8% of stroke codes (ischaemic stroke, haemorrhagic stroke, or unspecified stroke codes) appeared in the same hospital episode (ie. had the same event date, with one coded diagnosis in the primary position, and ≥1 other coded diagnosis in the secondary position).

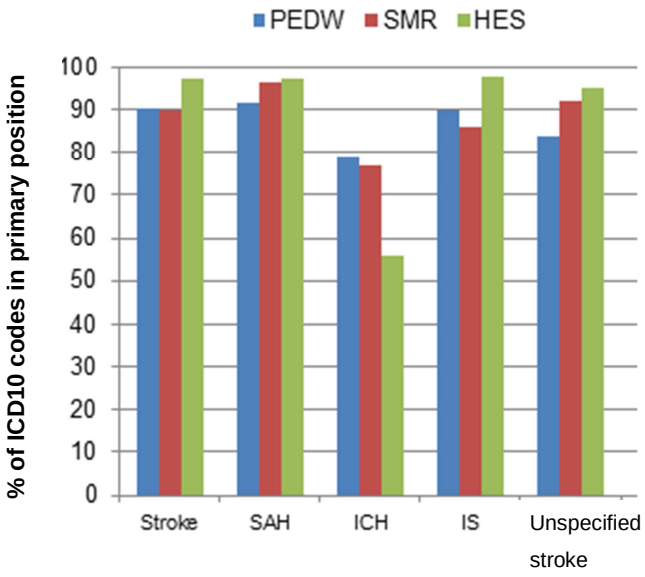

Figure 6.6 Proportions of stroke cases identified using hospital coded data (ICD-10 codes only) from England (HES), Scotland (SMR), and Wales (PEDW), which were in the primary diagnostic position, stratified by the main pathological type of stroke.

6.4.4 Identification of the main stroke types

Prevalent stroke

Amongst 502, 523 participants in the whole UK Biobank cohort there were 11, 572 prevalent stroke cases identified by self-report and/or linkages to coded hospital data. The majority of these prevalent stroke cases had unspecified stroke codes (~78%), and only 22% had ischaemic or haemorrhagic stroke codes (~11% ischaemic stroke, ~7% SAH, ~5% ICH). When stratified by data source, cases identified in the coded hospital data were more often specified types of stroke (ischaemic or haemorrhagic) compared to cases reported by participants (~69% of 'coded' strokes were specified pathological types versus only ~7% of self-reported strokes). Furthermore, participants appeared more likely to report the type of stroke if it was haemorrhagic, rather than ischaemic. Amongst cases identified by participant self-report (n=11,572) ~0.2% were ischaemic stroke, ~5% were SAH, ~1.6% were ICH, and ~93% were unspecified type (Table 6.3).

Incident stroke

A higher proportion of incident stroke cases were specified as ischaemic or haemorrhagic stroke, compared to prevalent stroke cases (~89% versus ~22%, respectively). Amongst 961 incident stroke cases (where each participant could have more than one incident case, one of each pathological type), 55% were ischaemic stroke, 19% were SAH, 15% were ICH, and only 11% were unspecified type. Incident strokes of specified pathological type were more likely to be haemorrhagic stroke in coded death certificate data (~74% of 70 specified strokes), and more likely to be ischaemic stroke in coded hospital data (~97% of 751 specified strokes). Comparing coded hospital data before UKB recruitment to coded hospital data after UKB recruitment (and up to the end of December 2010), a higher proportion of ICD codes were specified ischaemic or haemorrhagic stroke after recruitment, than before recruitment (~31% versus ~17%, respectively).

Table 6.3 Proportions of prevalent and incident stroke cases in the whole UK Biobank cohort (n=502,523 participants) which were ischaemic, haemorrhagic, or unspecified stroke.

Coded data source	Participants* (n)	Stroke cases† (n)	IS‡ (% total)	SAH‡ (% total)	ICH‡ (% total)	Unspecified stroke‡ (% total)
Prevalent stroke						
Hospital§	2,786	3,135	1,257 (40)	596 (19)	311 (10)	971 (31)
Self-report	8,304	9,507	17 (0.2)	457 (5)	155 (1.6)	8,881 (93)
Combined	8,769	11,572	1,270 (11)	848 (7)	453 (4)	9,001 (78)
Incident stroke						
Hospital§	785	902	521(58)	97 (11)	133 (15)	151 (17)
Deaths§	95¶	105¶	18 (17)	26 (25)	26 (25)	35 (33)
Combined	850	961	531(55)	178 (19)	141 (15)	111 (11)

*The number of participants identified with prevalent or incident stroke.

†The number of stroke cases identified amongst participants with prevalent or incident stroke. The number of cases is greater than the number of participants, because some participants had more than one type of stroke (ie. an incident/prevalent ischaemic stroke and an incident/prevalent haemorrhagic stroke).

‡Number of cases specified as Ischaemic Stroke (IS), Subarachnoid haemorrhage (SAH), Intracerebral haemorrhage (ICH), or otherwise unspecified.

§Based on ICD codes I60, I61, I63, I64, 430, 431, 434, 436.

¶There were ten death certificates which included more than one stroke code (e.g . codes for more than one pathological type of stroke).

6.4.5 Characteristics of the denominator population

Characteristics of the whole UK Biobank cohort (n=502, 523) are displayed for comparison with characteristics of participants recruited in Scotland and Wales (n=56,663). Both populations were very similar (Table 6.4).

Table 6.4 Characteristics of the UK Biobank population (n=502, 523) and the population recruited in Scotland and Wales (n=56,663).

	Whole cohort No. (%)	**Scotland/Wales No. (%)**
Total no. of participants	502,523	56,663*
Age group		
≤ 44	51,812 (10.3)	5,745 (10.1)
45-49	66,085 (13.2)	7,892 (13.9)
50-54	76,335 (15.2)	9,137 (16.1)
55-59	90,826 (18.1)	10,837 (19.1)
60-64	121,478 (24.2)	12,728 (22.5)
65 +	95,987 (19.1)	10,324 (18.2)
Mean (std)	*56.5 (8.1)*	*56.3 (8.0)*
Gender		
Female	273,392 (54.4)	31,239 (55.1)
Male	229,131 (45.6)	25,424 (44.9)
Townsend deprivation index	-1.3 (3.1)	-1.3 (3.3)
Smoking status		
Never	273,545 (54.4)	31,279 (55.2)
Previous	173,066 (34.4)	18,357 (32.4)
Current	52,974 (10.5)	6,784 (12.0)
Unknown	2,938 (0.6)	243 (0.4)
Alcohol drinking status		
Never	22,538 (4.5)	2,372 (4.2)
Previous	18,112 (3.6)	2,166 (3.8)
Current	460,381 (91.6)	52,008 (91.8)
Unknown	1,492 (0.3)	117 (0.2)
Blood pressure, mean (std)		
SBP	139.8 (19.7)	140.9 (19.6)
DBP	82.2 (10.7)	83.2 (10.8)
BMI, mean (std)	27.4 (4.8)	27.6 (4.9)
Self-reported disease##		
Stroke	7,169 (1.4)	943 (1.7)
TIA	1,781 (0.3)	157 (0.3)
Other CVD	16,537 (3.3)	1,687 (3.0)
High cholesterol	61,625 (12.3)	6,368 (11.2)
Hypertension	133,290 (26.5)	15,342 (27.1)
Heart disease#	22,948 (4.6)	2,896 (5.1)
Renal diseases	884 (0.2)	116 (0.2)

#including angina, heart attack and heart failure;

##: self-reported diseases confirmed by verbal interview

*Our sub-cohort comprised 40% of these participants (n= 26,562). Subsequent analyses (performed after this thesis was written) confirm that characteristics of the sub-cohort are very similar to the original UK Biobank population.

6.5 Discussion

Using cohort wide linkages to multiple combinations of coded data increased the number of potential strokes identified amongst UK Biobank participants. No single data source identified all potential cases, and each different data source generated different estimates of stroke incidence and prevalence. The majority of prevalent stroke cases were identified by participant self-report (range ~90 to 94% of prevalent cases). The majority of incident stroke cases were identified using linkages to coded hospital data (range ~79 to 92% of incident cases). A large proportion of incident cases identified using hospital or death certificate coded data were specified types of stroke (up to 89% were ischaemic or haemorrhagic stroke). The addition of linked coded primary care data increased the overall numbers of stroke cases detected. Read codes identified ~39% of incident stroke cases and ~49% of these (~19% of incident cases) were not identified by any other data source. These additional cases would have been missed if hospital data and death certificate data (ICD codes) were the only source of stroke outcomes identification during UKB follow-up.

This study provides useful estimates of the contribution of each data source to the identification of stroke cases in UK Biobank. In future, potential cases will need to be validated using an independent gold standard to determine the accuracy of each data source (sensitivity and positive predictive value, PPV, for stroke and its main pathological types) in the UK Biobank population. I have previously shown that the PPVs of selected ICD codes for 'acute stroke', the codes used in this study, ranged from 68 to 90% in various populations worldwide (Chapter 2). PPVs were generally higher using primary position codes than secondary position codes, albeit by only 5 to 10% (Chapter 2). It is therefore encouraging that the majority of our ICD coded stroke diagnoses were identified in the primary diagnostic position (81% and 97%, respectively). It is also encouraging that a high proportion of our incident stroke cases were a specified type of stroke. In the past, ICD codes for ischaemic and haemorrhagic stroke (SAH or ICH) have performed consistently well, often with PPVs ≥90% (Chapter 2). Assuming that ICD codes perform equally well in the UKB population, linkages to these data will be sufficient to identify and classify a large proportion of our incident stroke cases.

The majority of our prevalent stroke cases were ascertained by self-report. In a previous chapter I demonstrated that the positive predictive value (PPV) of self-reported stroke was low in populations with low stroke prevalence (<10%). Under these circumstances, around 1/3 to 3/4 of self-reported strokes were false positives. (Chapter 3). In this chapter the estimated prevalence of stroke in the whole UK Biobank cohort was ~1.7% using a combination of self-report and ICD codes. In the UKB sub-cohort, stroke prevalence was ~2% based on self-report, Read codes, and ICD codes. In light of this low stroke prevalence, many of the potential stroke cases identified by participant self-report are likely to be false positives.

Self-report data generated different estimates of stroke prevalence in the UK Biobank sub-cohort compared to the whole UK Biobank population (4.6% versus 1.7%, respectively). The higher prevalence of self-reported stroke in the sub-cohort is potentially due to 'over- reporting' of stroke in this population, rather than a true difference in stroke prevalence. Published studies have suggested potential links between recent and/or frequent GP contact and over-reporting of certain medical conditions, including cardiovascular diseases, although no definite conclusions have been drawn.(Kreigsmann et al. 1996, Englert et al. 2010) The UK Biobank sub-cohort included only those individuals who had continuous registration with Scottish and Welsh GP practices providing data to UKB for linkage and analysis. It therefore excluded individuals who had not registered with a GP practice supplying data to UK Biobank. Although some of these excluded individuals had registered with GP practices not yet providing linked data to UK Biobank, others might not have registered with a GP. It is possible that individuals who attend their GP frequently are more aware of stroke/TIA as potential diagnoses, and are more likely to ascribe previous symptoms and signs to such a condition, compared to individuals who do not see their GP often. It is possible that the former group were overrepresented in the UK Biobank sub-cohort compared with the UK Biobank population as a whole. It is also possible that additional differences in characteristics between the UK Biobank sub-cohort and the whole UK Biobank population (e.g differences in age range(s) and or vascular risk-factors), may have led to a difference in stroke prevalence between these populations. Characteristics of the whole UK Biobank cohort and the UK Biobank Welsh and Scottish populations are very similar (Table

6.4), but I did not have access to characteristics of the ~22,000 sub-cohort participants. It is reassuring that when multiple overlapping data sources were used, the estimated stroke prevalence in these two groups were more similar (~1.7% in the whole UK Biobank cohort versus ~2% in the sub-cohort), compared to estimates based on self-report alone (~1.7% versus 4.6%, respectively). Although these additional data sources (coded hospital, death certificate and primary care data) are not considered 'gold standards' for stroke diagnosis, the similar estimates of stroke prevalence in both populations when these multiple sources agree, supports the idea that stroke prevalence was 'over-estimated' by self-report alone, particularly in the UK Biobank sub-cohort.

We have shown that large numbers of potential stroke cases were identified by participant self-report. If self-report were used as a screening tool in future, to increase the sensitivity of incident stroke detection during UKB follow-up, it is likely that this would generate a lot of work (ie. large numbers of potential stroke cases for subsequent validation, many of which may turn out to be false positive). Therefore, self-report is likely to be an inefficient method for the identification of stroke outcomes in UK Biobank.

For the purpose of nested case cohort or case control studies in UK Biobank, participants with prevalent stroke will be excluded to enable identification of participants with true 'first in a lifetime stroke' after their date of recruitment. In these studies it is important to exclude as many potential prevalent stroke cases at the outset as possible, to minimise the number of participants misclassified 'without stroke' at recruitment, and to reduce the risk of reverse causality bias. We can infer from this study, that self-report is likely to have high sensitivity for stroke, given the high proportion of cases identified using this method. Self-report is therefore likely to be a useful tool for the identification of 'stroke free' individuals at baseline, enabling these individuals to be included in nested case cohort or case control studies of the associations between risk factors and stroke.

Different forms of coded data (data from primary care, hospital admissions, and death certificates) identify different proportions of the main types and subtypes of stroke. In this study, the pathological types of stroke identified by death certificates

(ICD codes) were ~74% haemorrhagic stroke and ~26% ischaemic stroke, whereas the converse was true for hospital admissions (main types identified were ~69% ischaemic stroke and ~31% haemorrhagic stroke). This is in keeping with the fact that haemorrhagic stroke is more likely to be fatal than ischaemic stroke. Excluding any single data source from follow-up has the potential to under represent a particular type or subtype of stroke from the overall outcomes identified, and this may introduce bias in future associations between risk factors and stroke, its main types, and subtypes. Therefore, although relatively small proportions of incident strokes were identified using coded data from death certificates (~11% of incident stroke cases in the whole UKB population, and ~12% of incident stroke cases in the UKB sub-population), linkages to these data are important in order to avoid exclusion of fatal, non hospitalised stroke cases. The same principle applies to the use of primary care coded data for follow-up of stroke outcomes. It has been shown previously that non-fatal, non hospitalised strokes are more likely to be ischaemic stroke cases, and more likely to be due to small vessel disease. These cases may not be detected by any other data source. Using linkage to multiple coded data sources therefore improves the accuracy of stroke outcomes identification by ensuring that all types and subtypes of stroke are fairly represented.

Stroke prevalence in the whole UK Biobank population (~1%) is lower than published for the UK (~2.8% in 2010).(Townsend et al. 2012) Stroke incidence in UK Biobank was 850 per 502,523 from recruitment to end 2012, based on coded hospital and death certificate data (~0.09% per year, assuming average time from recruitment to end 2012 was ~2 years). UK Stroke incidence was ~165 cases per 100,000 in 2007, based on the same data sources (~0.17% per year).(Townsend et al. 2012). Stroke cases should occur less frequently in UK Biobank than in the general population due to inclusion of healthier than average participants ('the healthy cohort effect').(Lindsted et al. 1996) In addition, UK Biobank included younger participants than the population at large, meaning that stroke incidence and prevalence will be lower than average during early follow-up, but should increase with time (as stroke incidence increases with age).

6.6 Conclusions

In conclusion, linkages to multiple sources of coded data are likely to improve the accuracy and completeness of stroke outcomes identification. Potential cases will need to be validated using an independent gold standard to determine the PPV and/or sensitivity of these data sources for stroke and its main pathological types in UK Biobank. However, it is likely that a large proportion of incident strokes will be classified into main pathological types using coded data alone. If these coded data perform similarly in UK Biobank to other populations, this should be achieved with high accuracy, and further validation of these cases (in terms of their main pathological type) may not be necessary. Future work should explore the potential of coded primary care data and/or data extracted from hospital discharge summaries, to classify stroke into its main pathological subtypes.

Section C: Assessing the potential of coded primary care data to capture blood pressure variability in UK Biobank.

Chapter 7 Blood pressure variability and risk of stroke: a review of the evidence

- Blood pressure variability (BPV) may be an independent risk factor for stroke.
- It is unclear how BPV should be defined, measured, and quantified to predict stroke risk.
- Short-term, medium-term, and long-term BPV have different potential underlying causes. These different categories of BPV are only partly correlated, and some associate more strongly with cardiovascular risk than others.
- In this chapter I report my review summarising the most recent evidence for an association between medium-to-long term BPV and risk of stroke.
- I explored the influence of BPV measurement methods, adjustment (or not) for confounding factors, and participant characteristics, on the strength of the association between BPV and stroke.
- All included studies demonstrated a trend towards an association between BPV and stroke, but the majority were in highly selected populations.
- BPV appeared to increase with age, female sex, rising mean BP, and past history of vascular disease, but the association with stroke was stronger in younger age groups and in those with lower mean BP.
- The strength of the association between BPV and stroke may depend on the precision of BPV measurement
- Further large population based studies are required to reliably investigate the associations between BPV, stroke and its main pathological types.

7.1 Introduction

7.1.1 Blood pressure variability: a novel risk factor for stroke

Blood pressure (BP) is one of the most important modifiable risk factors for stroke. Average BP has a steep log-linear association with stroke, such that increases of 20mmHg systolic/10mmHg diastolic approximately double stroke risk in middle to

old age. Recent research suggests that blood pressure variability (BPV) is an independent risk factor for stroke.(Rothwell et al. 2010b) If found to be true, this would have important potential implications for BP monitoring and treatment, and may lead to new insights into stroke mechanistic pathways. In this chapter I have introduced the concept of blood pressure variability. I have explored how blood pressure variability is defined and quantified and have reviewed the recent evidence that increased BPV associates with stroke risk.

7.1.2 Treatment of blood pressure in current clinical practice

Blood pressure lowering drugs are among the most frequently prescribed medications in the UK.(Health and Social Care Information Centre 2012) Traditional approaches to BP monitoring and treatment have focussed on estimation of an individual's 'usual blood pressure'. 'Usual blood pressure' is the theoretical underlying value of blood pressure, which is widely considered to be the most important component of blood-pressure-associated vascular risk.(Rothwell 2010) In epidemiological studies the association between 'usual', or long-term average BP, and vascular risk is strengthened by adjusting for within-individual BPV over time (adjustment for regression dilution bias).(Clarke et al. 1999) BP variability has therefore traditionally been considered an obstacle to the estimation of an individual's 'true' underlying blood pressure.

In current clinical practice hypertension is not diagnosed unless blood pressure is consistently elevated. British Hypertension Society Guidelines (2011) recommend that if clinic BP readings are >140/90mmHg a patient is referred for Ambulatory Blood Pressure Monitoring (ABPM), or Home Blood Pressure Monitoring (HPBM), to get an accurate assessment of their 'usual blood pressure'.(http://www.nice.org.uk/guidance/cg127/chapter/1-recommendations) This recognises the fact that blood pressure is often raised in clinic compared to home readings (so called 'white coat hypertension') and avoids misdiagnosis of hypertension.

Results from the Health Survey for England (2006) showed that half of patients with hypertension remained ‘uncontrolled’ despite being prescribed anti-hypertensive medication.(Falaschetti et al. 2009) It has been shown that there is a tendency amongst primary care physicians to ‘ignore’ elevated clinic BP readings on the grounds that they are ‘not representative’ of a patient’s ‘usual’ BP level.(Nicodème et al. 2009) Non-compliance with prescribed medication is an important additional cause of poor BP control.(Wuerzner, Hassler and Burnier 2003, Munger, Van Tassell and LaFleur 2007) It is thought that ~ 40% patients who begin treatment for hypertension discontinue their treatment within 2 years.(MacDonald, Morant and Mozzafari 2007) In some cases discontinuation might be directed by the physician (particularly if subsequent BP readings are normal), but it is often because the patient experiences side effects, or fails to engage with their treatment.(Jin et al. 2008) Diagnosis and treatment of hypertension is challenging, and fluctuation of blood pressure from one clinic visit to the next certainly contributes to this challenge. What if identifying and treating BP variability is as important as the estimated underlying mean?

Part of improving BP control, and reducing cardiovascular risk, should focus on improving medication compliance, and on educating and empowering patients with high blood pressure to take charge of their own health. Patients may not understand why they are being treated for high blood pressure, particularly if some readings are normal, and others not. In this case there may be a tendency (amongst patients and their physicians), as suggested above, to focus on the ‘normal readings’ and ignore the elevated readings. If it were to be shown that variability itself increases vascular risk, and that this should be monitored and treated in addition to the mean level, it might encourage patients to engage more with their own risk stratification and treatment. Patients’ perception of their illness plays a role in non-adherence to prescribed anti-hypertensive medication,(Morrison et al. 2015) and compliance is generally improved when a patient believes in a therapy and perceives benefits from it.(Jin et al 2008). Therefore, demonstrating an association between BPV and risk of vascular disease, including stroke, would lead to an important change in the global approach to blood pressure control.

The key questions which follow are:

1: How do we define BP variability

2: At what point is BP fluctuation pathological, rather than normal?

3: How do we monitor BP variability, including over what time periods?

4: How do we risk-stratify, ie. how do we identify individuals at increased risk of adverse outcomes due to BPV?

7.1.3 Definitions of blood pressure variability

The British Society of Hypertension states that:

'Everybody's blood pressure is variable…further work is needed to define what is an acceptable level of variability and what might be considered extreme and thus worthy of further attention'.

Blood pressure variability can be divided into short-, medium-, and long-term variability (Table 7.1). These different categories of variability have different potential underlying causes, not all of which are 'pathological'. It has been shown that the different categories of BPV are only partly correlated, and that some associate more strongly with cardiovascular risk than others. For example, visit-to-visit BPV was more strongly associated with stroke and coronary events than both within-visit variability and variability during 24hr ambulatory blood pressure monitoring (ABPM) in post-hoc analyses of data from ~19,000 hypertensive individuals in the Anglo-Scandinavian Cardiac Outcomes Trial Blood Pressure Lowering Arm (ASCOT-BPLA).(Rothwell et al. 2010a)

Recognised forms of short-term variability (<24hr) include circadian variability (night-time dip or morning rise in blood pressure), and transient BP fluctuations due to stress, pain, emotion, changes in posture, or activity.(Parati et al. 2013, Rothwell 2010) White coat hypertension is a unique form of short-term variability, in which a patient's blood pressure is raised in clinic, but normal at home (or, alternatively, when the first clinic BP reading is high, but subsequent 'within-visit' readings are normal).(Celis and Fagard 2004) Short term variability is most often measured by 24 hour Ambulatory BP monitoring (24hr ABPM), or Home Blood Pressure Monitoring

(HBPM).(Table 7.1) Variability can also be captured by successive blood pressure readings within a single outpatient clinic visit ('within-visit variability'). Some forms of short term variability are normal physiological responses for example to physical activity, or pain, but excessive variability may be pathological (and may result, for example, from abnormal autonomic or humoral BP control, or increased arterial stiffness).(Rothwell 2010, Parati et al. 2013, Nagai and Kario 2013)

Table 7.1 Definitions and potential causes of blood pressure variability

Duration of variability	BP measurement method	Potential causes of variability
Short-term (<24 hours)	24hr ABPM[*] HBPM[†] Multiple 'within-visit' BP[‡]	Autonomic BP dysregulation: • increased sympathetic drive • reduced cardiopulmonary and arterial reflexes Reduced arterial compliance Psychological stress Behaviour (physical activity, posture, sleep). Endocrine factors (eg. insulin resistance)
Medium-term (≤ 30 days)	Multiple 24hr ABPM[*] HBPM[†] Visit-to-visit clinic BP[§]	Short-term BP variability Antihypertensive medication BP changes at work vs. weekend Inconsistent BP readings (measurement error)
Long-term (months/years)	As for medium-term	Short- and medium-term BP variability Seasonal BP variation Changes in mean BP over time Incident disease (eg. cardiovascular disease) affecting arterial compliance and autonomic BP regulation

[*]ABPM: 24 hour ambulatory blood pressure monitoring. This requires at least 2 BP measurements per hour during waking hours, and at least one BP measurement per hour overnight, over a 24 hour period. Blood pressure is measured in the patient's usual environment using an automated BP device

[†]HBPM: Blood pressure is measured in the patient's home, either 'self-measured', or taken by a health professional, at pre defined intervals. This may include morning and evening BP measurements to investigate diurnal variation in BP (short-term BPV), or morning/evening BP measurements over successive days (medium-term BPV).

[‡]BP readings taken over a period of minutes within a single outpatient clinic visit.

[§]BP readings taken at successive outpatient clinic visits.

Medium to long-term variability is due to changes in blood pressure over days, weeks, months, or years. Day-to-day variability can be captured using successive home BP readings, and week-to-week or month-to-month variability can be captured using successive outpatient clinic readings. Recognised forms of medium-to-long term variability include seasonal variation (BP is generally higher in winter than in summer),(Sega et al. 1998) and the gradual change in mean BP over time (mean BP tends to rise with age).(Drizd, Dannerberg and Engel 1986) Other factors which may contribute to medium-to-long term variability include medication (treatment with anti-hypertensive medication, particularly changes in treatment and/or non-compliance), and incident disease (for example BP instability may occur following stroke).(Rothwell 2010, Parati et al. 2013)

7.1.4 Biology of BPV and potential mechanisms of stroke causation

'Pathological' blood pressure variability is thought to result from a combination of increasing arterial stiffness due to age, atherosclerosis, diabetes, chronic kidney disease, and reduced baroreceptor function due to age, cerebral ischaemia, diabetes, hypertension, endothelial dysfunction, and 'inflammation'.(Rothwell 2010, Nagai and Kario 2013, Chapleau et al. 1995) In addition, what could be perceived as 'normal' physiological causes of short-term BPV include stress, emotion, pain, and posture, and the anticipated 'night time dip' and 'morning surge in blood pressure'. Other factors contributing to short or long-term BPV include medications (which modulate BP), and/or intercurrent illness, which may increase BP instability.

The mechanism by which BPV causes stroke is thought to be due to a combination of BP variability over time, and increased BP 'instability'- sudden peaks or troughs in BP - which together lead to cerebral hypoperfusion and ischaemia. In a normal

individual, autoregulation maintains cerebral perfusion over a range of blood pressures. Sudden peaks or troughs in BP are therefore unlikely to cause ischaemia unless there are other factors which render them vulnerable to BP instability, for example, extracranial arterial stenosis, or cerebral small vessel disease.

It is hypothesised that when baroreceptor dysfunction coexists with arterial stiffness, there is reduced ability to buffer short-term fluctuations in BP, and this leads to an overall increase in BP variation. In this context, fluctuations in blood pressure become sudden and larger, and may be enough to cause cerebral ischaemia even in the context of normal cerebral autoregulation. Furthermore, changes in mean BP over time may lead to a shift in the 'cerebral autoregulation curve', such that the thresholds of BP needed to induce ischaemia become different (for example, higher than normal in generally hypertensive individuals, and lower than normal in generally hypotensive individuals). Changes in mean BP over time, in combination with 'pathological' BP variation (above), are thought to increase an individual's risk of stroke by increasing the chance of cerebral hypoperfusion.(Rothwell 2010, Nagai and Kario 2013, Floras 2013). A similar mechanism is thought to underpin the potential association between increasing BPV and the development of brain white matter hyperintensities, and cognitive impairment.(Nagai and Kario 2013)

A different hypothesis is that blood pressure variability leads to stroke via endotheial dysfunction, inflammation and cerebral small vessel disease, but less has been published about this theory.(Rothwell 2010)

7.1.5 How is BPV quantified?

A number of different measures of blood pressure variability have been described. The simplest method is derived from an individual's maximum and minimum recorded blood pressure over a specified period of time. Individuals can be grouped into four categories which include: stable normotension, episodic moderate hypertension, episodic severe hypertension, and stable hypertension (Table 7.2). Alternatively, a continuous measurement can be derived by subtracting the minimum BP from the maximum BP recorded within the exposure period (MMD: maximum

minus minimum difference). Episodic hypertension and maximum blood pressure reached have been associated with risk of stroke (Rothwell 2010).

Table 7.2 Categorical definition of blood pressure variability

Category	BP readings (mmHg)
Stable normotension	Max. ≤ 140mmHg
Episodic moderate hypertension	Min. ≤140mmHg Max. 140-179mmHg
Episodic severe hypertension	Min. ≤ 140mmHg Max. ≥ 180mmHg
Stable hypertension	Min. ≥140mmHg

Repeated Systolic (SBP) or Diastolic (DBP) blood pressure readings (successive day-to-day, or visit-to-visit readings) can also be converted into continuous measures of blood pressure variability (Table 7.3). The more frequently used measures include standard deviation (SD), coefficient of variation (CV), variation independent of the mean (VIM), and average real variability (ARV). Standard deviation and coefficient of variation correlate with mean BP. 'Variation independent of the mean' (VIM) was developed to try and account for this co-linearity between SD and mean BP.(Rothwell et al. 2010b)

Amongst any group of BP readings, BP variability could result from visit-to-visit variability or from a shift in the underlying BP over time. Figure 7.1 illustrates how SD BP fails to distinguish between two scenarios, the first (Patient A) where there is visit-to-visit variability about a stable mean BP, and the second (Patient B), where there is development of stable normotension in a previously stable hypertensive patient. By contrast, Average Real Variability (ARV), the average absolute difference between successive BP values specifically captures BP variability from one visit to the next. (Mena et al. 2005) ARV SBP is therefore the best measure of time-series variability (variability from one reading to the next). If SD, CV and VIM

are used to capture BP variability, further adjustments for changes in mean BP over time may need to be considered. This is particularly relevant if BP variability is being measured over longer exposure periods where changes in the mean BP might be expected (eg. seasonal variation, an increase in BP with age, or a reduction in BP with treatment).

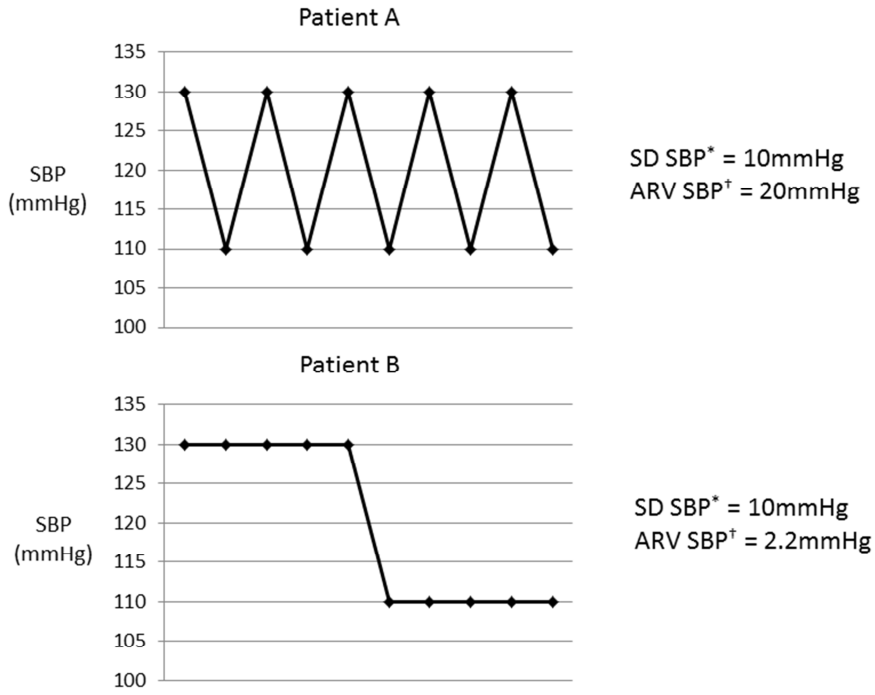

Figure 7.1 An illustration of the difference between SD SBP and ARV SBP using ten consecutive BP readings from two patients with very different patterns of BPV.[‡]

[*]SD SBP is the same for Patient A and Patient B despite a very different pattern of BP variability. SD SBP captures overall variability, irrespective of whether this is variability from one visit to the next, or a change in the underlying BP over time.

[†]ARV SBP is higher in Patient A, where visit-to-visit variability is greater, and lower in Patient B, where visit-to-visit variability is lower. By contrast, SD SBP is the same for both patients. ARV SBP is a more specific measure of visit-to-visit BP variability compared to SD SBP.

[‡]Adapted from Mena et al., 'A reliable index for the prognostic significance of blood pressure variability.' Journal of Hypertension 2005; 23:505-511

Another measure of BP variability, which is less affected by changes in mean BP over time, is the ‘SDreg’ (Standard deviation regression). Whereas SD in SBP is the 'average' of the deviations about the mean BP (which is assumed to be static over time), SDreg is the 'average' of the deviations about the regression line (which is the ‘line of best fit’ for a linear increase in mean BP over time). Therefore in a participant whose BP does not change across visits, SD and SDreg are similar. In contrast, in a participant whose BP increases linearly over time, SD is higher than SDreg, and SDreg more accurately captures variability independent of the change in underlying mean BP.(Shimbo et al. 2012)

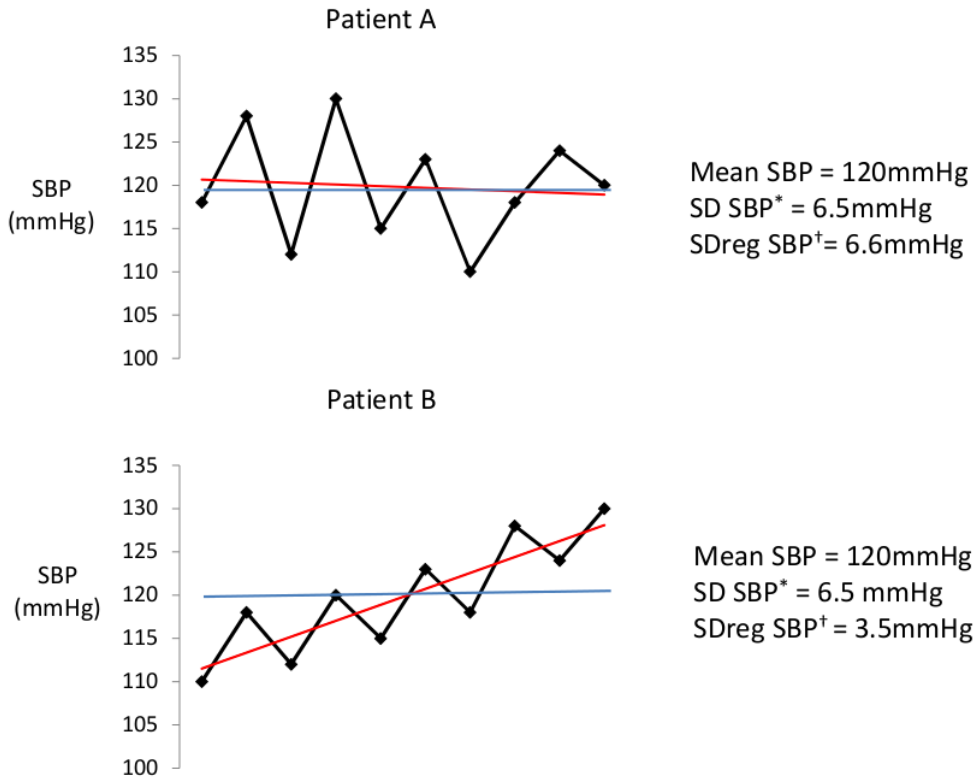

Figure 7.2‡ An illustration of SD SBP and SDreg SBP using ten consecutive readings from two different patients , one with and one without a change in underlying BP over time.

*SD BP calculates variability about the mean BP over time (blue line).

†SDreg BP captures variability about the change in BP over time (red line). This is considered an ‘adjustment’ for the change in mean BP over time.

Patient A: SD SBP and SDreg SBP are similar, because there is little change in the underlying BP over time.

Patient B: SDreg SBP is lower than SD SBP because there is a gradual increase in the underlying BP over time. SDreg therefore more accurately reflects that the visit-to-visit variability is lower for Patient B than for Patient A.

‡Adapted from Shimbo et al. 'Association Between Annual Visit-to-Visit Blood Pressure Variability and Stroke in postmenopausal women: data from the Women's Health Initiative.' Hypertension 2012:60:625-630.

Table 7.3 Quantifying medium to long term BPV

BPV measurement	Calculation (successive home or visit-to-visit BP values)
1. Standard Deviation (SD)	$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$
2. Coefficient of Variation (CV)	SD/mean
3. Variation Independent of Mean (VIM)	$SD/mean^x$ (x derived from curve fitting)
4. Average Real Variability (ARV)	Average absolute difference between successive BP values
5. Standard Deviation 'regression' (SDreg)	Variation about the linear increase in mean BP over time.

7.1.6 What evidence is there that BPV associates with risk of stroke?

There is currently a lack of consensus as to whether blood pressure variability increases risk of stroke.(Floras 2013) A post hoc analysis of blood pressure control using data from the INVEST trial, 2007 (patients with hypertension randomised to

either verapamil or B-blocker based therapy) assessed whether vascular outcomes were related to consistency of BP control (% of time BP was <140/90). Risk of stroke (but not MI) decreased according to the proportion of visits where BP was controlled (analyses adjusted for baseline BP, and on-treatment mean BP). This suggested that mean BP does not provide the complete picture of BP related risk.(Mancia et al. 2007) More compelling evidence for a potential association between blood pressure variability and stroke, again based on post hoc analyses of trial data, was published in 2010.(Rothwell et al. 2010b) Visit-to-visit blood pressure variability was a strong predictor of stroke amongst a population with recent TIA (Top versus bottom decile HR for risk of stroke 6.22, 95% CI 4.16-9.29, adjusted for mean BP). These results were replicated in a number of other TIA cohorts.(Rothwell et al. 2010b) However, subsequent analyses have not shown consistent associations.(Floras 2013) The most recently published study, a meta-analysis of 2 prospective studies (60,096 patients from US and Europe), found a marginal association between BPV and risk of stroke (HR 1.02, 95% CI 1.01-1.03 per 1mmHg increase SD SBP),(Tai et al. 2015) but this study did not include all available, relevant, published data.

7.1.7 Investigation of BPV and risk of stroke in UKB

In order to examine the association between BPV and risk of stroke in UK Biobank, I need to know if there is any evidence to suggest how analyses between BPV and risk of stroke should be performed, including how an individual's BP should be measured, and how BPV should be quantified. For example, is there any consensus on the number of BP readings required, duration of measurement (days versus months versus years), gaps between successive measures (days versus weeks versus months), calculation of BPV (SD, versus CV, versus VIM, versus ARV), and adjustments for other factors (eg. comorbidities, anti-hypertensive medication, mean BP), which best predict risk of stroke? With this in mind, I performed a review of the literature, to summarise the most recent evidence for an association between medium-to-long term BPV and risk of stroke, and to investigate how BPV has been defined and quantified.

7.2 Methods

7.2.1 Search strategy and inclusion criteria

I searched MEDLINE, EMBASE and Google Scholar until the end of June 2015 using search terms ‘blood pressure variability’ and ‘stroke’. I searched for studies of the association between any continuous measure of medium to long-term BPV (eg. SD, CV, VIM, ARV, or transformations of these measures) and risk of stroke in any human adult population. I excluded studies which measured intermediate disease outcomes, (eg. carotid-intima media thickness, ‘silent infarction’, or white matter hyper-intensities)and excluded studies in post stroke populations because of the risk of reverse causality (increased blood pressure variability immediately following stroke). I also excluded studies which only measured short-term variability (24hr BP variability, or increased diurnal variability), because these measures appear to associate less strongly with cardiovascular outcomes than medium to long term variability.(Rothwell 2010) Furthermore, data on 24hr ABPM are not currently available in UK Biobank or any other very large prospective study. Medium-term blood pressure variability was defined as ‘day to day’ variability, and long-term blood pressure variability was defined as ‘month-to-month’, or ’year-to-year’ variability. I reviewed titles and abstracts of the first 300 results of each search, sorted by relevance, and full-texts of any potentially relevant articles. Finally, I reviewed bibliographies of included publications to search for additional relevant articles.

7.2.2 Data extraction

I extracted data from each included study on first author, publication year, population characteristics (country, age range, comorbidities), BP measurement (number of BP readings taken, time period over which BP measurements were repeated, time between repeated BP readings, if specified), stroke outcome measured and numbers (all stroke, fatal or non fatal stroke, ischaemic or haemorrhagic stroke), method of outcomes ascertainment (eg. expert medical record review, routinely coded data, or a combination of methods), duration of follow-up, BP variability measure(s) used, confounding factors accounted for in analyses (eg. mean BP, age, sex, cardiovascular risk factors), and HR (95% CI) for stroke.

Where a number of different statistical models were used, I extracted the result with the largest number of potential confounding factors accounted for (including mean BP at baseline and/or change in mean BP over time). I chose to report results for variability in systolic rather than diastolic BP because systolic BP (mean or variability) is a stronger predictor of stroke than diastolic BP (mean or variability).(Rothwell 2010) Finally, I extracted data from included studies on participant characteristics associated with increased BPV, and on any characteristics (from subgroup analyses) which increased the BPV associated risk of stroke.

7.2.3 Data analysis

I tabulated results for visual inspection. I did not perform meta-analysis or meta-regression because of heterogeneity between included studies in population characteristics, BP measurement methods (number of readings, duration of measurement, and gap between measurements), and reporting of results.
Visit-to-visit variability in blood pressure increases with the number of BP readings, the time period over which successive readings are taken, and the gap between successive readings.(Levitan et al. 2012) These factors need to be considered when comparing studies, and may introduce bias (if differences between subjects in numbers of BP readings, and/or period over which readings are taken, are independently associated with factors which increase the risk of stroke). Where possible, I performed within study comparisons to investigate the influence of the BPV measure used on the risk of stroke (including number of BP readings, and/or time period over which readings were taken, and/or the statistical method used to quantify BPV). I also compared the influence of adjustment or not for confounding factors (including change in mean BP over time, and/or treatment with anti-hypertensive medication) on prediction of risk. Finally, I looked for any evidence of differences (within each study) in the strength of associations between BPV and the main stroke types (ischaemic versus haemorrhagic stroke).

7.3 Results

7.3.1 Results of search

From 951 titles and abstracts, I reviewed 75 full-texts, and included 16 publications. (Figure 7.3).

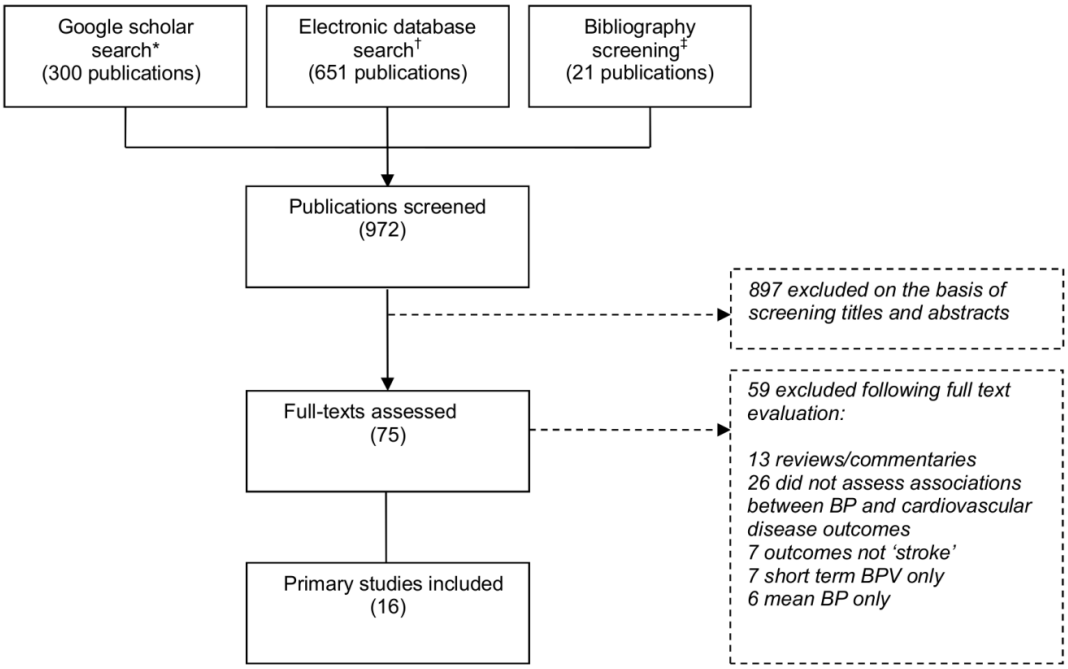

Figure 0.1 Flow diagram showing selection of primary studies which investigated associations between medium to long term BP variability and risk of stroke.

*First 300 publications sorted by relevance out of 369,000.

†MEDLINE and EMBASE search using the terms 'Blood pressure variability' and 'stroke'.

‡Bibliographies of included publications were searched for additional potentially relevant studies.

Twelve of the excluded full-texts were reviews/commentaries and one was a systematic review and meta-analysis of data from two already included publications.(Tai et al. 2015) Seven studies were excluded because they assessed the composite outcomes of cardiovascular disease and/or all cause mortality instead of stroke, or because they assessed intermediate disease outcomes (eg. brain infarction) instead of stroke, (Johansson et al. 2012, Muntner et al. 2011b, Rossignol et al. 2012, Mancia 2012, Hsieh et al. 2012, Eguchi et al. 2012, Brickman et al. 2010) six

because they assessed short-term rather than medium- or long-term BP variability,(Pierdomenico et al. 2009, Eto et al. 2005, Meredith et al. 1995, Pringle et al. 2003, Frattola et al. 1993, Mena et al. 2005) seven because they assessed associations between mean BP (not variability) and risk of stroke,(Verdecchia et al. 2001, Myint et al. 2008, Rapsomaniki et al. 2014, Lewington et al. 2002, Lim et al. 2012, MacMahon et al. 1990, Ohkubo et al. 2004) four because they studied reproducibility in BPV measurements over time but not association of BPV with stroke,(Kimberlin and Winterstein 2008, Altman and Bland 1986, Muntner et al. 2011a, Howard and Rothwell 2009) nine because they examined the relative effectiveness of different measures of BPV (ie. HBPM vs ABPM vs clinic BP), (Hodgkinson et al. 2011, Padfield 2010, Sheppard et al. 2014, Hodgkinson et al. 2014, McManus et al. 2014, Warren et al. 2010, Dolan et al. 2005, Mann, Millar Craig and Raftery 1985, Mancia et al. 2012) and fourteen for a variety of reasons, including that they compared statistical measures of BPV but did not assess association between BP and outcomes (including stroke), or because they examined factors influencing BP variability.

7.3.2 Characteristics of included studies

Population characteristics

Results are displayed in Table 7.4. Sixteen studies assessed the association between medium-to-long-term BP variability and risk of stroke in eighteen different populations. Six of these studies were post hoc analyses of data from clinical trials, nine were prospective observational studies, and one was a case control study.(Hata et al. 2000) Populations were from Europe (four UK based), US, Australia, Japan, China, Bangladesh, or multiple countries worldwide. Populations ranged in age from >18 to >70 years. Six of the populations studied were patients with hypertension,(Rothwell et al. 2010b, Carr et al. 2012, Chowdhury et al. 2014, Hata et al. 2000, Hastie et al. 2013, Yu et al. 2014) four had recent TIA,(Rothwell et al. 2010b) one had type-2 diabetes,(Hata et al. 2013) and one had ‘vascular disease’ (including cardiovascular disease, type 2 diabetes, and cardiovascular risk factors).(Poortvliet et al. 2012) One study included men only,(Hashimoto et al. 2012) and another included women on dietary modification/hormone therapy.(Shimbo et al. 2012) The remaining studies were performed in relatively unselected

populations.(Schutte et al. 2012, Asayama et al. 2013, Kikuya et al. 2008, Yinon et al. 2013, Suchy-Dicey et al. 2013, Gao et al. 2014)

The range of individuals with a diagnosis of hypertension was 24% to 71% amongst studies which did not select hypertensive participants.(Gao et al. 2014, Poortvliet et al. 2012, Hata et al. 2013, Kikuya et al. 2008, Schutte et al. 2012) From 2% to 39% of these participants were on anti-hypertensive medication at baseline.(Rothwell et al. 2010b, Yinon et al. 2013, Hashimoto et al 2012, Kikuya et al. 2012, Schutte et al. 2012, Asayama et al. 2013). Amongst studies which selected for a diagnosis of hypertension, the proportion of participants on anti-hypertensive medication at baseline ranged from 50% to 100%.(Carr et al. 2012, Hata et al. 2000, Chowdhury et al. 2014). In general, the proportion of participants with a diagnosis of hypertension at baseline and those ‘on’ versus ‘not on’ anti-hypertensive medication was incompletely reported.

Table 0.1 Results from sixteen included studies of the association between medium- to long-term systolic BPV and risk of stroke.

First Author	Country	Population (n)	Selection criteria	Age (range)	BP readings (n, duration)	Gap between readings	Follow-up (mean years)	Outcomes* (n)	BPV measure†	HR stroke (95%CI)
Schutte 2012	Belgium	2,944	Population based	45‡	2 home readings§ (28 days)	2-4 weeks	12¶	Fatal stroke (49)	VIM**	**1.13 (0.88-1.46)** 1 SD increase VIM SBP
									ARV**	**1.10 (0.81-1.49)** 1 SD increase ARV SBP
									MMD**	**1.17 (0.92- 1.48)** 1 SD increase MMD SBP
Asayama†† 2013	Japan	~2,400‡‡	Population based	35-96	26 self-measured¶ (≥5 days)	24 hours	11.9	All stroke (223)	VIM**	**1.14 (1.00-1.30)** 1 SD increase VIM SBP
									ARV**	**1.13 (1.00-1.27)** 1 SD increase ARV SBP
									MMD**	**1.12 (0.98-1.28)** 1 SD increase MMD SBP
Kikuya†† 2008	Japan	~2,400‡‡	Population based	35-96	26 self-measured¶ (≥10 days)	24 hours	12¶	Fatal stroke (83)	SD**	**1.38 (1.12-1.72)** 1 SD increase SD SBP
								Fatal IS (51)		**1.46 (1.11-1.91)** 1 SD increase SD SBP
Hashimoto 2012	Japan	902	Men without stroke	35-92	≥10 self-measured (≥10 days)	24 hours	13.1¶	All stroke (123)	SD**	**1.15 (0.96-1.36)** 3mmHg increase SD SBP
								IS (89)		**1.26 (1.02-1.55)** 3mmHg increase SD SBP

Hashimoto 2012	Japan	902	Men without stroke	35-92	≥10 self-measured (≥10 days)	24 hours	13.1[¶]	ICH (28)		**0.97 (0.65-1.46)** 3mmHg increase SD SBP
								Fatal stroke (38)		**1.47 (1.11-1.45)** 3mmHg increase SD SBP
Yinon 2013	Bangladesh	11,153	Population based	≥18		2.2 years		Fatal stroke[§§] (66)	SD[¶¶]	**1.51 (0.93-2.44)** 1 SD increase in log SBP
Suchy-Dicey 2013	US	1,642	Population based	≥65	5 clinic readings (5 years)	1 year	9.9	All stroke (195)	SD[§§§]	**1.03 (0.89-1.21)** 1 SD increase in variability
		1,028	Hypertension					All stroke -	SD	**1.00 (0.97-1.03)** 1 SD increase in SD SBP
		1,115	Hypertension					All stroke -	SD	**1.06 (0.92-1.22)** 1 SD increase in SD SBP
Hata 2013	Worldwide[***]	8,811	Type 2 diabetes	≥55	6 trial readings (21 months)	1-6 months	2.4[¶]	All stroke (176)	SD[†††]	**1.08 (0.93-1.25)** 1 SD increase SD SBP
									CV[†††]	**1.08 (0.93-1.25)** 1 SD increase CV SBP
Shimbo 2012	US	58,228	Women post menopause	50-70	8 trial readings[‡] (8 years)[‡]	1 year	5.4[¶]	All stroke (997)	SD[**]	**1.72 (1.28-2.32)** Top vs. bottom quartile SD SBP
									SDreg[**]	**1.46 (1.15-1.85)** Top vs. bottom quartile SD SBP
									SD[**]	**1.16 (1.08-1.24)** 5mmHg increase SD SBP
									SDreg[**]	**1.12 (1.05-1.19)** 5mmHg increase SD SBP

Rothwell 2010	UK	1,324	TIA	60‡	7 trial readings (2.5 years)	4 months	~3¶	All stroke (104)	SD**,‡‡‡	**4.84 (3.03-7.74)** Top vs. bottom decile SD SBP
									CV**,‡‡‡	**4.61 (3.11-6.83)** Top vs. bottom decile CV SBP
									VIM**,‡‡‡	**3.88 (2.13-5.38)** Top vs. bottom decile VIM SBP
	UK/Ireland Scandinavia	1,012	TIA + hypertension on atenolol	40-79	~ 13 trial readings (5.5 years)	1-6 months	5.5¶	All stroke -	SD**,‡‡‡	**4.29 (1.78-10.36)** Top vs. bottom decile SD SBP
									CV**,‡‡‡	**3.51 (1.56-7.53)** Top vs. bottom decile CV SBP
									VIM **,‡‡‡	**3.96 (1.66-9.43)** Top vs. bottom decile VIM SBP
	UK/Ireland Scandinavia	999	TIA + hypertension on amlodipine	40-79	~ 13 trial readings (5.5 years)	1-6 months	5.5¶	All stroke -	SD**,‡‡‡	**4.39 (1.68-11.50)** Top vs. bottom decile SD SBP
									CV**,‡‡‡	**3.25 (1.32-8.00)** Top vs. bottom decile CV SBP
									VIM**,‡‡‡	**3.57 (1.38-9.19)** Top vs. bottom decile VIM SBP
	Europe	1,247	TIA	≥18	8 trial readings (2 years)	3 months	2	All stroke -	SD**,‡‡‡	**1.78 (1.21-2.62)** Top vs. bottom decile SD SBP
									CV**,‡‡‡	**2.22 (1.52-3.22)** Top vs. bottom decile CV SBP
									VIM**,‡‡‡	**1.86 (1.28-2.69)** Top vs. bottom decile VIM SBP

Rothwell 2010	Netherlands	3,150	TIA on aspirin	-	7 trial readings (~ 2.5 years)	4 months	2.6	All stroke -	SD[**,‡‡‡]	**3.35 (1.63-6.87)** Top vs. bottom decile SD SBP
									CV[**,‡‡‡]	**3.41 (1.62-7.19)** Top vs. bottom decile CV SBP
									VIM[**,‡‡‡]	**1.83 (0.76-4.39)** Top vs. bottom decile VIM SBP
Carr 2012	UK	4,396	hypertension	65-74	~20 trial readings ≥5 years	2-12 weeks	5.9	All stroke -	Other[§§§]	**1.15 (1.01-1.31)** 1SD increase in BPV
Chowdhury 2014	Australia	5,880	hypertension	≥65	8 trial readings[‡] -	6 months	4.1[¶]	All stroke	SD[**]	**2.78 (1.28-6.05)** Top vs. bottom decile SD SBP
Poortvliet 2012	UK/ Netherlands[††]	4,819	vascular disease	70-82	5 trial readings (1 year)	3 months	3	All stroke (158)	SD[**,§§§]	**1.2 (0.8-1.9)** Top vs bottom quartile SD SBP
	UK[††]	1,805	vascular disease	70-82	9 trial readings (2 years)	3 months	7.1	All stroke (245)	SD[**,§§§]	**1.3 (0.9-1.80)** Top vs. bottom quartile SD SBP
Hata 2000[¶¶¶]	Japan	488	hypertension	≥60	~ 12 clinic readings 1 year	~ 1 month	-[¶¶¶]	IS (138)	CV[****]	**OR 1.15 (1.03-1.29)** 2% increase in CV SBP[****]
Hastie 2013	UK	7,136	hypertension[††]	20-80	4 clinic readings[‡] (1 year)	≥30 days	1[††††]	Fatal stroke (318)	ARV[‡‡‡‡]	**1.18 (0.83-1.69)** Top vs. bottom quartile ARV SBP
		5,715	hypertension[††]		7 clinic readings[‡] (2-5 years)		2-5[††††]	Fatal stroke (210)		**1.34 (0.80-2.22)** Top vs. bottom quartile ARV SBP
		2,685	hypertension[††]		6 clinic readings[‡] (5-10 years)		5-10[††††]	Fatal stroke (80)		**1.26 (0.63-2.5)** Top vs. bottom quartile ARV SBP

Hastie 2013	UK	1,292	hypertension[††]	20-80	- (>10 years)	≥30 days	>10[††††]	Fatal stroke (33)		**0.931 (0.33-1.63)** Top vs. bottom quartile ARV SBP
Gao 2014	Japan	2,906	No cognitive impairment	≥60	≥6 clinic readings (≥2 years)	-	12.9[¶]	Fatal CVD[§§§§] (123)	Other[§§§]	**0.63 (0.28-0.44)** Low vs. high variability
Yu 2014	China	122,636	hypertension	≥18	≥ 6 clinic readings (~3 years)	≤6 months	4	CVD[¶¶¶¶] (4,522)	SD[**]	**1.93 (1.74-2.15)** Top vs. bottom quartile SD SBP

Shaded results are from post hoc analyses of data from clinical trials.

[*]Stroke outcomes determined by medical record review, unless otherwise specified. CVD= Cerebrovascular Disease.

[†]Blood Pressure Variability measure(s):VIM=Variation Independent of the mean, SD= Standard Deviation, CV= Coefficient of Variation, ARV= Average Real Variability, MMD= Maximum minus minimum blood pressure difference, SDreg= Standard deviation regression.

HR= Hazard Ratio. SBP= Systolic Blood Pressure

[‡]Mean

[§]Ten nurse measured BP readings taken on two separate occasions.

[¶]Median

[**]Adjusted for mean BP, age, sex, anti-hypertensive medication use (where relevant), and other cardiovascular risk factors.

[††]These data are from the same population.

[‡‡]Only 48-49% of original population, because participants who were not at home/hospitalised, did not take home BP readings, or had <5 BP readings were excluded.

[§§] I60-I69 ICD-10 codes for stroke confirmed by verbal autopsy.

[¶¶]Adjusted for age, sex, baseline SBP, heart rate, and betel leaf use.

[***] 20 countries from Asia, Australasia, Europe, and North America.

[†††]Adjusted for age, sex, anti-hypertensive medication, glucose control intervention, cardiovascular risk factors, mean BP during measurement period.

[‡‡‡]Also adjusted for change in mean BP over time and results were similar.

[§§§]SD in SBP adjusted for change in mean BP over time.

¶¶¶Case control study which used BP readings 1 year prior to development of stroke (and readings over 1 year for age matched controls).

****Adjusted for age, sex, presence/absence AF, frequency of hospital visits, proteinuria, smoking, status and mean BP. Found a similar results when patients with diagnosed AF were excluded.

††††This is a time to event analysis, so BPV is measured until outcome (stroke) or death.

‡‡‡‡Adjusted for age, sex, smoking, alcohol, BMI, cholesterol, prevalent CVD and estimated glomerular filtration rate (eGFR).

§§§§Based on death certificate codes ICD-9: 430-438 and ICD-10: I60-I69

¶¶¶¶Based on ICD-10 codes I63-I67, I60 and G45.

BPV measurement methods

Included studies varied widely in their BP measurement methods. Blood pressure was measured in different environments (home vs. clinic), at different times of day (morning vs. evening vs. unspecified), using different observers (self-measured vs. clinician-measured), using different BP devices (manual vs. automated vs. unspecified), and with different measurement protocols (some published, some not). Some studies used a single BP reading per clinic visit (or a single BP reading per day, if HBPM was used), whereas others used the mean of multiple 'within-visit' or 'within-day' readings to get a single BP reading per day or per visit.

Blood pressure variability (BPV) was estimated over days to weeks (medium-term) or over months to years (long-term). The number of separate 'visits' (home or clinic BP readings) used to derive visit-to-visit BPV ranged from 2 to 26, and the gap between these visits ranged from 24 hours to 2.2 years. One study which estimated BPV using only 2 visits included 5 separate BP readings from within each visit.(Schutte et al. 2012) The remaining studies required a minimum of 3 separate visits to estimate visit-to-visit BPV.

Studies performed in trial cohorts used BP readings from ~5 to 20 outpatient visits taken over ~1 to 8 years. The time between successive visits was often standardised in clinical trials (mean gap amongst included studies ~4 months), and BP measurement followed published protocols. BPV measurement (number of readings, duration, and between visit gap) varied more widely amongst prospective observational studies. At one extreme, three studies used home BP measurements taken over successive days for a period of ~1 month,(Asayama et al. 2013, Hashimoto et al. 2012, Kikuya et al. 2012) whereas at the other extreme one study used biennial BP readings taken over a period of 6 years.(Yinon et al. 2013)

Four studies used routinely recorded BP measurements to estimate BPV.(Hata et al. 2000, Hastie et al. 2013, Gao et al. 2014, Yu et al. 2014) Blood pressure data were retrospectively extracted from hypertension outpatient clinics (Hata et al. 2000, Hastie et al. 2013), or from primary care medical records (Gao et al. 2014, Yu et al. 2014). Three of these studies were in hypertensive populations (Hata et al. 2000, Hastie et al. 2013, Yu et al. 2014) and in the fourth ~ 70% of included individuals

were hypertensive.(Gao et al. 2014) The number of BP readings and between-visit gaps were not standardised in these studies.

BPV was derived using a number of different statistical approaches. The main methods used (and reported) were SD, VIM, ARV, or a transformation in which SD SBP was adjusted for changes in mean BP over time.

Follow-up and outcomes ascertainment

Participants were followed up for 2 to ~13 years. In general, clinical trial populations were followed-up for a shorter duration (range 2 to 7 years) than prospective observational studies (range 1 to 13 years). Where reported, numbers of outcomes were generally small (38 to 318) for stroke or fatal stroke, and smaller for pathological types of stroke (28 to 89), apart from one US study in women (~1,000 stroke outcomes).(Shimbo et al. 2012) The largest study population was ~12,600 Chinese hypertensive patients which yielded 4,522 cerebrovascular events (combined stroke, TIA and other cerebrovascular disease).(Yu et al. 2014) Stroke diagnoses were mostly adjudicated by physician review of medical record data, or by expert committees. Two studies used routinely coded data for stroke outcome identification (International Classification of Diseases codes, ICD, for stroke, or cerebrovascular disease). (Gao et al. 2014, Yu et al. 2014)

7.3.3 Overall results for BPV and stroke risk

All included studies demonstrated at least a trend towards an association between increasing BPV and increased risk of stroke. Results were significant in seven studies (amongst these studies the hazard ratio, HR, for stroke ranged from 1.15 to 4.84 for higher versus lower BPV),(Kikuya et al. 2012, Shimbo et al. 2012, Rothwell et al. 2010b, Carr et al. 2012, Chowdhury et al. 2014, Hata et al. 2000, Yu et al. 2014) but did not reach significance, or were of marginal significance, in the remaining nine (amongst these studies the HR for stroke ranged from 1.03 to 1.51 for higher versus lower BPV).(Schutte et al. 2012, Asayama et al. 2013, Hashimoto et al. 2012, Yinon et al. 2013, Suchey-Dicey et al. 2013, Hata et al. 2013, Poortvliet et al. 2012, Hastie et al. 2013)

A statistically significant (or marginally significant) association between BPV and risk of stroke was detected in relatively unselected populations,(Asayama et al. 2013, Kikuya et al. 2012, Shimbo et al. 2012), as well as in populations with cerebrovascular disease or hypertension.(Rothwell et al. 2010b, Carr et al. 2012, Chowdhury et al. 2014, Hata et al. 2000, Yu et al. 2014). Associations were also detected when BPV was estimated over shorter (≤30 days),(Asayama et al. 2013, Kikuya et al. 2012) as well as over much longer time periods (years).(Shimbo et al. 2012, Rothwell et al. 2010b, Carr et al. 2012, Yu et al. 2014)

All but one of the included studies (Hastie et al. 2013, which used ARV to measure BPV) included mean BP as a covariate in Cox regression analyses. All of the included studies also adjusted for age, and a selection of cardiovascular risk factors (Table 7.4). Three studies adjusted for renal disease (presence/absence of proteinuria, estimated Glomerular Filtration Rate, eGFR) (Hastie et al. 2013, Hata et al. 2000, Yu et al. 2014). Ten studies adjusted for anti-hypertensive medication use.(Yu et al. 2014, Poortvliet et al. 2012, Chowdhury et al. 2014, Rothwell et al. 2010b, Shimbo et al. 2012, Hata et al. 2013, Hashimoto et al. 2012, Kikuya et al. 2012, Asayama et al. 2013, Schutte et al. 2012) Of the nine studies which estimated BPV over ≥ 2 years, (Yu et al. 2014, Gao et al. 2014, Hastie et al. 2013 Poortvliet et al. 2012, Carr et al. 2012, Rothwell et al. 2010b, Shimbo et al. 2012, Suchey-Dicey et al. 2013, Yinon et al. 2013), six adjusted analyses for change in mean BP over time.(Gao et al. 2014, Poortvliet et al. 2012, Carr et al. 2012, Rothwell et al. 2010b, Suchy-Dicey et al. 2013, Shimbo et al. 2012) Of those which did not consider such adjustment, one used average real variability as a measure of BPV.(Hastie et al. 2013)

Influence of BPV measurement method(s) on stroke risk

Rothwell showed that the predictive power of BPV may depend on the precision of BPV measurement.(Rothwell et al. 2010b) When the number of BP readings used to estimate BPV increased from 4 to 10 amongst 2,006 individuals in the UK-TIA cohort (nb.the time period over which measurements were taken also increased from ~7 to ~ 36 months), the HR for stroke comparing top vs. bottom quintile SD SBP increased from 1.51 (95% CI 0.86-2.66) to 13.04 (95% CI 1.66-102.6). A similar pattern, but smaller effect, was demonstrated in the Women's Health Initiative

trial.(Shimbo et al. 2012) Using 3 to 4 BP readings (over ~ 4 years), the HR for stroke with 5mmHG increase SDreg SBP was 1.02 (0.91-1.14), and using 10 to 11 BP readings (over ~11 years) the HR for stroke with 5mmHg increase in SDreg SBP increased to 1.44 (0.85-2.45). In studies of medium-term BPV and stroke risk (ie. BP measured over days rather than over months to years), the number of BP readings (and duration over which readings were taken) had a marginal or no effect on stroke risk.(Asayama et al. 2013, Kikuya et al. 2012) Comparing 5 readings over 5 days versus ~26 readings over ~26 days the HR for stroke were 1.13 (95% CI 0.99-1.28) and 1.14 (95% CI 1.00-1.30), respectively.(Asayama et al. 2013) Similar results were found comparing BP readings over 10 days to readings over ~26 days (Second study in the same population).(Kikuya et al. 2012)

Influence of statistical analysis method on stroke risk

Five studies published results for BPV and risk of stroke using different statistical measures of BPV including SD, CV, VIM, ARV, and maximum minus minimum blood pressure difference.(Rothwell et al. 2010b, Schutte et al. 2012, Asayama et al. 2013, Suchy-Dicey et al. 2013, Hata et al. 2013) All five studies found similar results regardless of the method used to quantify BPV. Two additional studies reported that results were similar for SD SBP and CV SBP, and chose to report only SD SBP.(Kikuya et al. 2012, Poortvliet et al. 2012)

Rothwell adjusted analyses for change in BP over time as well as for mean BP, and this adjustment did not alter the risk associations, however results were not published. Shimbo compared HR stroke with and without adjustment for change in mean BP over time by comparing HR stroke using SDreg SBP with HR stroke using SD SBP (BPV measured over mean duration 8 years).(Shimbo et al. 2012) Risk associations were slightly attenuated ‘with’ compared to ‘without' adjustment for change in mean BP over time’, HR for top versus bottom quartile SDreg SBP 1.46 (95% CI 1.15-1.85) versus HR for top versus bottom quartile SD SBP 1.72 (95% CI 1.82-2.32). Shimbo reported results from the same population treating BPV as a continuous rather than a categorical variable. The magnitude of effect was smaller when BPV was treated as a continuous variable, HR stroke with 5mmHg increase in SD SBP was 1.16 (95% CI 1.08-1.24) and HR stroke with 5mmHg increase in SDreg

SBP was 1.12 (95% CI 1.05-1.19), compared to when BPV was treated as a categorical variable (comparing top versus bottom quartile BPV, above).

Blood pressure variability and the risk of main stroke types.

Blood pressure variability was associated with a trend towards an increased risk of both ischaemic and haemorrhagic stroke in the Women's Health Initiative trial (Shimbo et al. 2012), but the magnitude of effect was small in both cases, and 95% confidence intervals crossed zero for haemorrhagie stroke. HR for 5mmHg increase in SDreg SBP was 1.12 (95% CI 1.04-1.21) for ischaemic stroke, and 1.14 (95% CI 0.97-1.34) for haemorrhagic stroke. In two other studies,(Kikuya et al. 2012, Rothwell et al. 2010b) BPV was associated with increased risk of ischaemic, but not haemorrhagic stroke, but smaller numbers of haemorrhagic vs. ischaemic stroke outcomes may have made it harder to detect an association in the haemorrhagic subgroup. In one Japanese study HR for 3mmHg increase SD SDP was 1.26 (95% CI 1.02-1.55) for ischaemic stroke and 0.97 (95% CI 0.65-1.46) for intracranial haemorrhage, based on 89 ischaemic and 29 haemorrhagic stroke outcomes.(Kikuya et al. 2012)

Characteristics associated with BPV and with increased BPV-related stroke risk

Amongst included studies, blood pressure variability increased with age,(Rothwell et al. 2010b, Schutte et al. 2012, Hata et al. 2013, Shimbo et al. 2012, Hastie et al. 2013) female sex,(Hata et al. 2013, Rothwell et al. 2010b. Hastie et al. 2013) mean blood pressure,(Shimbo et al. 2012, Rothwell et al. 2010b, Kikuya et al. 2012, Hata et al. 2013, Hastie et al. 2013) and past history of vascular disease, including baseline chronic kidney disease.(Rothwell et al. 2010b, Hata et al. 2013, Hastie et al. 2013) However, the association between BPV and risk of stroke was stronger in younger age groups, (Rothwell et al. 2010b, Shimbo et al. 2012) and amongst those with lower mean BP.(Rothwell et al. 2010b, Shimbo et al. 2012)

Influence of anti-hypertensive medication on BPV and BPV-related stroke risk

Four studies reported that antihypertensive medication use was associated with higher indices of blood pressure variability,(Asayama et al. 2013, Suchy-Dicey et al.

2013, Hata et al. 2013, Shimbo et al. 2012) one found no such association,(Hata et al. 2000) and three reported that B-blockers were associated with increased BPV compared to other classes of anti-hypertensive medication.(Schutte et al. 2012, Rothwell et al. 2010b, Carr et al. 2012) Three studies reported analyses for the association between BPV and risk of stroke in subgroups of patients 'on' versus patients 'not on' anti-hypertensive medication.(Rothwell et al. 2010b, Suchy-Dicey et al. 2013, Shimbo et al. 2012) Rothwell found that visit-to-visit BPV predicted risk of stroke irrespective of treatment, HR for top versus bottom decile VIM SBP 3.67 (95% CI 2.34-5.75) for those 'on' treatment, and 2.27 (95% CI 1.41-3.67) for those 'not on' treatment.(Rothwell et al. 2010b) Suchy-Dicey found no difference in the association between BPV and risk of stroke amongst patients without hypertension and 'not on treatment'(HR 1.03 for 1SD increase SD SBP, 95% CI 0.89-1.21), compared to those with hypertension on 'stable treatment' (HR 1.00, 95% CI 0.97-1.03), or compared to those with hypertension and 'stopping/starting' treatment (HR 1.06, 95% CI 0.92-1.22).(Suchy-Dicey et al. 2013) Shimbo found that increased BPV was significantly associated with risk of stroke in ~33,000 individuals 'not on' antihypertensive treatment (HR for stroke associated with a 5mmHg increase SDreg SBP was 1.21, 95% CI 1.06-1.37), but not in ~13,000 individuals 'on' antihypertensive treatment (HR 1.06, 95% CI 0.97-1.16). However, the association was positive in both groups, and smaller numbers of participants 'on treatment' may have reduced power to detect a significant association in this group.(Shimbo et al. 2012).

7.4 Discussion

7.4.1 Summary of findings

All included studies demonstrated a trend towards an association between increased medium to long-term BPV and risk of stroke. Hazard ratios for stroke ranged from 1.03 to 1.51 for a 1 standard deviation increase in BPV (8 studies), from 0.93 to 1.93 for top versus bottom quartiles of BPV (4 studies) and from 1.78 to 4.84 for top versus bottom deciles of BPV (2 studies). One Japanese study showed that lower BPV reduced the risk of fatal cerebrovascular disease (HR 0.63, 95% CI 0.28-0.44 for low versus high BPV). In the only case control study, the odds ratio for stroke associated with a 2% increase in CV BP was 1.15 (95% CI 1.03-1.29).

Included studies varied widely in their BP measurement methods, approaches to quantifying BPV, duration of follow up (and numbers of stroke outcomes detected), statistical analyses (particularly approaches to adjustment for mean BP), and reporting of results. In a number of studies, blood pressure variability increased with age, female sex, rising mean BP, and past history of vascular disease. In two studies, the association between BPV and stroke was stronger in younger age groups and in those with lower mean BPV.

BPV is an important potential risk factor for stroke, but there is a lack of consensus on how BPV should be measured to best predict stroke risk. This is the biggest challenge to investigating associations between BPV and stroke risk in future studies, and needs to be overcome in the long-term if BPV is to become a target for stroke prevention in clinical practice.

7.4.2 Limitations of previous studies

Included studies have had a number of limitations. First, small numbers of outcomes (33 to 318 strokes in all but one study) have limited statistical power to detect the association between BPV and risk of stroke reliably. The study with the largest number of stroke outcomes (n=997) did show a statistically significant association between increased BPV and increased risk of stroke,(Shimbo et al. 2012) but these data were from a clinical trial and BP measurement methods (including setting, use of standardised measurement protocols, and the relatively fixed gap between

successive ‘visits’) are therefore less generalizable to the ‘real world setting’. Second, many of the included studies were in selected populations (amongst individuals with TIA, hypertension, diabetes or other vascular risk factors), and results of these studies are less generalizable to the population as a whole, particularly if the association between BPV and stroke is stronger amongst individuals with already elevated vascular risk. Of note, all UK based studies selected individuals with hypertension. Fourth, only three studies used BP data captured in routine clinical practice to estimate BPV. The remaining studies measured BP for research purposes (clinical trials, or prospective observational studies which measured BPV for a fixed duration, at relatively fixed intervals, and following published protocols). BP measurements in these studies are therefore not representative of ‘real world’ clinical practice. Of the three studies which used ‘routinely captured data’, two were in hypertensive populations, two analysed CVD outcomes (not stroke), and two used outpatient BP readings (rather than primary care data). The only study which used routinely captured primary care data to measure BPV was based in Japan, and this looked at fatal CVD outcomes (based on ICD coded data only). Sixth, it is not yet clear if differences exist in the strength of the association between BPV and the main stroke types (ischaemic vs. haemorrhagic stroke), and larger studies, accruing more outcomes overall, will be required to test these associations. Seventh, there have been no studies of the associations between BPV and the risk of ischaemic or haemorrhagic stroke subtypes.

7.4.3 Strengths of the review

This review includes more of the available, relevant data than previous published reviews of BPV and risk of stroke/cardiovascular disease.(Tai et al. 2015) Additional strengths include firstly the focus on stroke outcomes, rather than all cerebrovascular disease, or all cardiovascular disease. Some studies have shown differences in the strength of associations between BPV and risk of all cause mortality versus cardiovascular disease, versus stroke, versus MI.(Kikuya et al. 2012, Yinon et al. 2013, Suchy-Dicey et al. 2013, Hata et al. 2013) Secondly, I chose to focus on medium-to-long term BPV, rather than all forms of BPV, because BPV over these different durations (<24hr versus day-to-day versus month-to-month) may represent different clinical or pathological phenomena.(Rothwell 2010, Parati et al. 2013)

Amongst included studies, I have shown that an association between BPV and stroke risk exists in relatively unselected populations, as well as in those with increased vascular risk (eg. TIA/hypertension cohorts), and that 'real world data' (eg. BP measurements taken in routine primary care or outpatient clinics) can be used to predict risk of stroke as well as more standardised data (eg. BP readings from clinical trials). Finally, I have shown (in a number of different within study comparisons) that the association between medium-to-long term BPV and stroke risk did not appear to be affected by the statistical measure of BPV, or by the presence or absence of adjustment for mean BP over time.

7.4.4 Limitations of this review

Although I included more of the relevant published data for stroke than previous similar reviews,(Tai et al. 2015, Floras 2013, Mancia 2012, Nagai and Kario 2013, Diaz et al. 2014, Hocht 2013) I did not assess for inclusion all titles and abstracts from my initial search, and I may have inadvertently excluded some relevant articles. Furthermore, the trend towards an association between BPV and stroke may be due to publication bias (but this seems unlikely, given the current lack of consensus amongst experts as to whether BPV is a true independent risk factor for stroke). The association between increased BPV and risk of stroke was significant in some but not all included studies. This might reflect the relatively small numbers of strokes detected in some studies, reduced precision of BPV measurement (which may be a function of number of measures, duration of measurement, and gap between measurements), and different methods of reporting results (for example, comparing top versus bottom deciles or quartiles of BPV and stroke risk), or it may mean that increased BPV does not associate with increased risk of stroke in all circumstances. Included studies were from a wide variety of countries and populations, and differences in genetics, age, vascular risk factors, and/or the distribution of main stroke types (haemorrhagic stroke being more common in Asia than in Western Europe) may influence the overall association between BPV and risk of stroke. It was not possible to reconcile the above differences between studies, and for this reason it was also not possible to perform meta-analysis or meta-regression. Previous research has shown that visit-to-visit BP variability increases with the number of BP readings, with increased gap between BP readings (and over longer time periods),

and with single visit measures (vs. mean of multiple measures).(Levitan et al. 2012) It is important to consider such differences in these between studies, particularly if the number of BP readings taken per individual (or study) is correlated with an independent risk factor for stroke. Eight of the included studies did not use standardised numbers of BP readings amongst included participants, and therefore risk this form of bias.

Included studies adjusted for different confounding factors. It is possible that the association detected between BPV and risk of stroke is the result of residual confounding in some studies. One potential factor, not considered in many studies, was the presence/absence of renal disease. Chronic kidney disease is independently associated with both BPV,(Nakano et al. 2015) and with stroke,(Masson et al. 2015) and may account for some of the BPV associated stroke risk. However, three of the included studies which did adjust for some form of renal disease,(Hastie et al. 2013, Hata et al. 2000, Yu et al. 2014) found a marginal association, or non-significant trend towards an association, between BPV and risk of stroke.

Some critics of the BPV hypothesis suggest that BPV is due to an effect of anti-hypertensive medication (potentially variable compliance, or non-compliance) and that BPV associated vascular risk might be entirely avoided by better adherence to established BP treatment guidelines. However, one study included in this review found similar associations between BPV and risk of stroke in subgroup analyses of patients off anti-hypertensive medication, compared to those on anti-hypertensive medication, or those starting/stopping anti-hypertensive medication.(Suchy-Dicey et al. 2013) Other studies found similar associations between BPV and risk of stroke irrespective of treatment or not, whilst some found stronger associations between BPV and risk of stroke in those off treatment. The argument that BPV is a definite physiological and pathological phenomenon is further confirmed by the fact that BPV is reproducible within individuals over time.(Muntner et al. 2011a, Howard and Rothwell 2009, Rothwell et al. 2010b)

Other critics suggest that BPV associated vascular risk is small compared to mean BP, so that even if BPV is an independent risk factor for stroke, most of the risk can still be attributed to mean BP, and therefore this should remain the focus of clinical

risk assessment and treatment. Subgroup analyses (amongst studies included in this review) have shown that BPV associated stroke risk is higher in those with lower mean BP, and is also increased in younger versus older age groups (age groups which conventionally have less hypertension and less competing vascular risk factors overall). Therefore, although BPV may be more common amongst those with high blood pressure, and although BPV contributes a greater proportion of risk (relative to mean BP) amongst those with high BP, it may still be important to identify and treat BPV amongst younger individuals with normal blood pressure (for risk prevention). Furthermore, in the young, strokes are often cryptogenic, so it is particularly important to consider and investigate new risk factors in this group of individuals.

7.5 Conclusions and future directions

There is definite evidence that BPV is an important potential risk factor for stroke. We need to continue to explore which groups of individuals have increased BPV, and which are at most risk of stroke from increased BPV (and therefore would benefit most from new approaches to BP monitoring and treatment). It appears possible to demonstrate an association between BPV and risk of stroke using routinely captured 'real world' BP data, but previous studies which used these methods have had small numbers of stroke outcomes, have been in selected populations, or have used less precise methods of stroke confirmation (based on ICD-coded data only). As yet, there has been no study of the association between BPV and risk of stroke using data from a relatively unselected UK based primary care population. It is also not yet clear if the association between BPV and risk of stroke varies by stroke type (ischaemic versus haemorrhagic stroke), and there has been no study of the association between BPV and ischaemic and haemorrhagic subtypes. UK Biobank has the potential to overcome these limitations, due to large numbers of anticipated stroke outcomes, access to linked coded primary care, hospital and mortality data, and development of detailed stroke outcomes ascertainment and confirmation methods (outlined in previous chapters).

7.5.1 Future aims

In the following chapter I aim to explore the potential of routinely captured, coded primary care BP data to identify BPV in the UK-Biobank population. If coded data

can be used to identify individuals with increased BPV, it may be used to measure BPV in future studies of associations between BPV and risk of stroke, its main types and subtypes, in UK Biobank.

Chapter 8 Can coded primary care data be used to measure blood pressure variability?

- Blood pressure variability (BPV) may be an independent risk factor for stroke, but further large population based studies are required to reliably investigate this association.
- Confirming an association between BPV and stroke has important potential implications for BP monitoring and treatment and may provide new insights into stroke mechanistic pathways.
- UK Biobank has linkages for follow-up to national coded healthcare data. These data might be used to capture a novel exposure, such as BPV.
- In this chapter I report my investigation of the potential of using routinely collected coded primary care data to estimate visit-to-visit BPV in a sub-cohort of ~10,000 UK Biobank participants.
- I identified participants in whom visit-to-visit BPV could be measured using coded systolic blood pressure values (BP), explored the range of variability detected (SD BPV), the generalisability of populations selected, and the potential accuracy of coded BP data.
- I found that visit-to-visit BPV was captured in the majority of participants. Selecting participants with more coded BP data reduced generalizability, but there was good variability in BPV amongst those selected, and reasonable agreement between coded BP and an independent reference standard.
- I concluded that routinely collected coded primary care data could estimate BPV in UK Biobank participants. This would enable future exploration of associations between BPV and stroke, in a 'real world' setting, and in larger numbers of individuals than previously possible.

8.1 Introduction

8.1.1 Aims

In this chapter, I explored the feasibility of using routinely captured, coded primary care data to measure blood pressure variability in UK Biobank participants. I used de-identified data from >10,000 Welsh UK Biobank participants for whom linked coded primary care data are available through the Secure Anonymised Information Linkage (SAIL) platform hosted by Swansea University. In the long run, this work will inform potential approaches to BPV measurement in UK Biobank, enabling future nested case cohort or case control studies of the associations between BPV, stroke, its main types and subtypes.

8.1.2 Exploring associations between BPV and stroke in UKB

In the previous chapter, I showed that there is a trend towards an association between blood pressure variability and risk of stroke. However, previous studies have had a number of limitations including use of data from restricted populations eg. clinical trial populations and, or populations with TIA, hypertension, diabetes, or other risk factors, and small numbers of stroke outcomes, some of which were based on less accurate sources of data (eg. ICD codes only).

UK Biobank (UKB) has the potential to overcome some of these limitations. Around 5,000 incident stroke cases are expected to have occurred amongst UK Biobank participants by the end of 2017 (~4,000 ischaemic and ~1,000 haemorrhagic). These anticipated outcomes are around ten times the numbers available in previous studies (Chapter 7). Furthermore, UK Biobank is developing more accurate methods of stroke outcomes confirmation and sub-classification than used in previous, large, prospective observational studies. Therefore, if BPV could be measured in a sufficient proportion of UKB participants, it would enable large numbers to be included in future studies of the potential associations between BPV, stroke, and its main types in UKB.

8.1.3 Measuring BPV: the potential of coded primary care data

UK Biobank has cohort wide linkages to routinely collected coded primary care data. Although these data were primarily linked for participant follow-up and outcomes adjudication, they could also be used to measure exposures. It may be possible to use primary care coded data (BP related Read codes and the coded BP values associated with them) to measure visit-to-visit BPV.

The advantage of using coded primary care data to measure additional exposures (those not measured at the baseline assessment) is that it is an efficient and cost effective method because these data have already been collected. This is particularly relevant for the measurement of visit-to-visit blood pressure variability (BPV), which would require each participant to undergo multiple BP measurements at multiple different outpatient visits. Using previously collected data means that a larger number of UKB participants would be included in any future study than would otherwise be possible. This would make the most of the large numbers of anticipated stroke outcomes (above). Furthermore, because coded primary care data are collected in routine clinical practice, it would enable exploration of the association between BPV and stroke in a 'real world' setting.

8.1.4 BP in the coded primary care record: the Quality Outcomes Framework (QOF).

Using primary care data to estimate within participant BPV requires a sufficient number of individuals to have a sufficient number of coded BP measurements in their primary care record. Since its inception in 2004, the UK primary care Quality Outcomes Framework (QOF) has included targets for the proportion of individuals aged 45 years and older, registered with a GP, who have had a BP measurement. The quality outcomes framework (QOF) became part of the general practitioner contract in April 2004. Participation is voluntary, but all practices in Wales participated. The QOF rewards individual GP practices for the provision of 'quality care' by helping to fund the delivery of care. Payment is contingent on a practice achieving pre specified targets. For example in 2004-2005, 55 to 75% of those 45 years and older in the UK

were required to have at least one blood pressure reading recorded within the previous 5 years. The proportion was increased in 2006-2007 to include at least 65 to 80% of individuals over the age of 45. Stricter targets are in place for BP measurements amongst individuals with various vascular risk factors including diabetes, known hypertension, cardiovascular disease (stroke, TIA and coronary heart disease), and renal disease. From 2004, these individuals were required to have a BP measurement recorded in their medical records every 15 months (or every 9 months for those with established hypertension). Since 2011, these targets have also applied to individuals with schizophrenia, bipolar disease, other psychoses, peripheral arterial disease, and those with rheumatoid arthritis, and from 2013 onwards, annual BP readings have been required.(http://www.wales.nhs.uk/sites3/page.cfm?orgid=480&pid=10486) With these incentives in place, and with the fact that Welsh GP practices all participate in QOF, a large proportion of individuals should have at least annual or biennial BP readings recorded in the UK Biobank Welsh primary care dataset. Furthermore, I hypothesised that the number and frequency of BP readings recorded in the coded primary care record is likely to have increased over time (partly due to these incentives, and partly due to the changeover from handwritten to computerised medical records).

8.1.5 Objectives

My specific objectives for this chapter were as follows:

1) To determine what proportion of the UKB Welsh population have sufficient coded primary care data to estimate visit-to-visit BPV.

2) To determine the numbers of participants who can be identified with ‘x’ visits (each with coded BP readings), ‘y’ times prior to UK Biobank recruitment.

3) To determine generalisability of individuals with coded BP values to the whole population (UKB Welsh participants with coded primary care data)

4) To determine the variability in BPV amongst selected individuals, to judge if this is sufficient to be able to detect an association between increasing BPV and increasing risk of stroke.

5) To compare within individual agreement between the mean BP recorded at the UKB baseline assessment, and the mean BP estimated using coded primary care data.

8.2 Methods

8.2.1 Defining the denominator population

For this work I used de-identified coded primary care data from 9,947 UK Biobank (UKB) participants recruited in Wales between 2006 and 2010. Coded primary care data were first made available to UK Biobank in early 2014 through the Secure Anonymised Information Linkage (SAIL) platform hosted by Swansea University.(Ford et al. 2009, Lyons et al. 2009) At the time of this study, SAIL received data from ~ 50% of Welsh GP practices, and provided linked coded primary care data for follow-up of 11,147 Welsh UK Biobank participants (~50% of the 21,000 UKB participants recruited in Wales).

I requested de-identified coded primary care data from these 11,147 UK Biobank participants in early 2014. I downloaded all available coded data (version 2 Read codes from categories A to Z, 0 to 9, or from the 'drugs and appliances dictionary') into *stata* version 12 at the end of June 2014.

From the original population (n=11, 147), I excluded 1,200 who did not have 'continuous registration' with a 'SAIL GP practice' (a Welsh GP practice which provided coded primary care data to SAIL). I defined continuous registration as having a gap of <30 days between successive SAIL GP practice registrations. Participants were therefore excluded if they were registered for >30 days outside of Wales, or if they spent >30 days in Wales registered with a non SAIL GP practice.

Data provided until 31/1/2012 were assumed to be 'complete' for all SAIL GP practices. I excluded 8,658 records entered after this date (0.2% of all records) because some SAIL GP practices might not have uploaded their most recent data. I also excluded 7,786 coded events which had an 'impossible' event date, ie. the event date was before the participant's date of birth. The final population with continuous registration in a Welsh SAIL GP practice until the end of 2012 was 9,947. These individuals represented ~47% of Welsh UKB participants, and ~2% of the original UKB cohort.

8.2.2 Primary care data structure and definitions

The basic structure of the primary care coded dataset (provided in the SAIL database) is displayed in Table 8.1. BP related Read codes are shown for illustration, but the original dataset included all coded events (Read codes from any chapter plus drug prescriptions).

***Table 8.1 Structure of the primary care coded dataset*[*]**

PT_ID_E	DOB	EVENT_CD	TERM_60	EVENT_DT	VALUE1	VALUE2
1[†]	D/M/Y	662..	Hypertension monitoring[§]	D/M/Y[‡]	0[§]	.[§]
1[†]	D/M/Y	246..	O/E - BP reading	D/M/Y[‡]	120	80
1[†]	D/M/Y	246..	O/E - BP reading	D/M/Y	110	70
2	D/M/Y	2469.	O/E - systolic BP[¶]	D/M/Y	140	.
3	D/M/Y	G2…	Hypertensive disease[§]	D/M/Y	0[§]	.[§]
3	D/M/Y	246..	O/E – BP reading	D/M/Y	0	100
4	D/M/Y	246A.	O/E - diastolic BP[¶]	D/M/Y	100	.

PT_ID_E: Anonymous participant identifier

DOB: date of birth

EVENT_CD: Read code

TERM_60: Text definition accompanying the Read code

EVENT_DT: Event date, which is the presumed date of the GP visit.

VALUE1: Usually systolic BP (for a BP-related Read code).

VALUE2: Usually diastolic BP (for a BP-related Read code).

[*]BP related read codes are used as an example, however the original dataset included the full available primary care coded record for each participant, and therefore included all types of Read codes (categories A-Z, 0 to 9, and medication codes) .

[†]Each new row of data is a new record (a new Read code with its associated participant identifier, text definition, event date, +/- values). I grouped records by participant identifier, and then sorted them in chronological order.

[‡]If two different records had the same participant identifier and same event date, they were assumed to have originated from the same GP visit. Multiple Read codes can be entered to describe different aspects of a single GP visit (eg. symptom codes, examination findings, administration codes, prescription codes).

[§]Not all BP related Read codes had associated BP values. ‘0’ and ‘.’ were ‘missing values’.

¶Twenty two Read codes specified the type of BP measurement in the Read code text definition, eg. systolic BP or diastolic BP. These codes made up <1% of all BP codes with values. For the remainder, VALUE1 was assumed to be systolic BP, and VALUE2 was assumed to be diastolic BP.

I defined various terms, as follows:

1) Coded event: each new row of data. Each coded event included an anonymous participant identifier (and date of birth), a Read code with its accompanying text definition, and an event date. The event date was the presumed date of the GP visit (and the closest approximation to the date on which the event occurred). Some (but not all) coded events with BP related Read codes included BP values (VALUE1 and/or VALUE2).

2) A participant's coded primary care record: All coded events with the same, unique, participant identifier made up that participant's coded primary care record (represented by shading in Table 8.1).

3) GP visit: For each participant, coded events with the same event date were assumed to have arisen from the same GP visit. Multiple coded events (eg. multiple Read codes +/- multiple coded BP values) could have been recorded at a single GP visit.

8.2.3 Characteristics of the denominator population

I counted the number of coded events included in this dataset, and the number of coded events included per individual. I grouped records by year to count the numbers of events available from 1980 to the end of 2013 (note, I did this step before censoring records at the end of 2012, section 8.2.1, above). I investigated the duration of coded events, per individual, using the last coded event date – first coded event date. I then investigated characteristics of the denominator population including age, gender (% male), and % of vascular risk factors using linkage to the UK Biobank baseline assessment dataset (see section 8.2.8, below, for methods).

8.2.4 Creating Read code lists for BP

My original dataset included Read codes for every possible diagnosis. However, I aimed to identify groups of individuals who had blood pressure related Read codes in their coded primary care record. This would enable me to identify individuals who had 'probable' or 'possible' hypertension, as well as individuals (with or without hypertension) who had their BP measured. The first step in this process was creating a 'Read code list' for blood pressure.

I used three sources to search for Read codes related to blood pressure (specifically, 5-byte version 2 Read codes, as this was the version used in SAIL at the time of my study). Firstly, I searched the NHS Read code browser using the terms 'blood pressure', 'hypertension', 'hypotension', 'BP', 'hypert', 'blood-pressure', 'labile', 'high', and 'low'. I then manually searched the browser for additional Read codes, looking within the following categories/subcategories: 'family history', 'cardiovascular symptoms', 'past medical history', 'history of cardiovascular disease, 'suspected condition', 'examination of cardiovascular system', 'on examination: blood pressure reading', 'hypertensive disease', and 'cardiovascular drugs'. Secondly, I included Read codes from the 'QOF hypertension indicator set for Wales' which I found on the NHS Wales website (http://www.wales.nhs.uk/), using the search terms 'QOF', 'Read codes' and 'hypertension'. Thirdly, I searched two CALIBER primary care code lists for blood pressure: hypertension diagnosis, and 'blood pressure category'.(http://www.caliberresearch.org/)

I included Read codes from all three sources, (ie. any Read code potentially related to blood pressure) apart from Read codes for cardiovascular drugs (eg. thiazide diuretics, calcium channel blockers, loop diuretics etc.), because these can be prescribed for a number of different medical conditions and are therefore less specific to 'hypertension'.

8.2.5 Identifying participants with 'probable' or 'possible' hypertension based on Read code text definitions.

After identifying as many BP related Read codes as possible, I divided codes into five categories based on their text definitions (see Table 8.1 for examples of 'text definitions'). The five categories were:

1) Hypertension ‘diagnosis’ codes: diagnostic codes which identified participants with ‘probable’ hypertension (eg., ‘G….’ for ‘hypertensive disease’).

2) Hypertension ‘administration’ codes: administrative codes which identified participants with ‘possible’ hypertension (eg., ‘9N03.’ for ‘seen in hypertension clinic’)

3) Blood pressure ‘monitoring’ codes: codes which implied that blood pressure was measured, some of which also identified participants with ‘possible’ hypertension (eg., ‘662..’ for ‘hypertension monitoring’).

4) Hypotension ‘diagnosis’ codes: diagnostic codes which identified participants with hypotension (eg., ‘G870.’ for ‘orthostatic hypotension’)

5) Blood pressure ‘value’ codes: codes which implied that they were associated with recorded BP values (eg., ‘246..’ for ‘on examination: blood pressure reading’)

I searched for each Read code individually to see how frequently it was recorded in the dataset. I then identified the number of participants (out of 9,947 in my denominator population) who had any Read code related to BP. From within this population I searched for individuals who had a ‘probable’ or ‘possible’ diagnosis of hypertension using combinations of Read codes from category 1 (above, for ‘hypertension diagnosis’) and selected Read codes from categories 2 to 3 (above, for ‘hypertension administration’, or ‘blood pressure monitoring’). See Appendix 8.6 (Table 8.14) for the groups of Read codes used to define ‘probable’ or ‘possible’ hypertension. Finally, from within the population identified with ‘probable’ hypertension, I searched for numbers of participants who had at least one coded event for ‘blood pressure monitoring’ and/or at least one coded event with ‘blood pressure values’. I repeated this for the population of individuals without any ‘probable hypertension’ Read codes. I constructed a Venn diagram to display the proportion of individuals with and without ‘probable’ hypertension who had at least one instance of blood pressure monitoring, and/or at least one recorded BP value. Analyses were performed in *Stata* (version 12).

8.2.6 Creating a dataset of systolic blood pressure (SBP) values.

I identified the proportion of BP Read codes in the dataset, stratified by category (1 to 5, above), which had associated BP values. I expected blood pressure 'monitoring' and blood pressure 'value' codes to have associated BP values, whereas I did not expect hypertension 'diagnosis' or 'administration' Read codes to have BP values.

If 'monitoring' or 'value' Read codes did not have BP values (either VALUE1 or VALUE2 present), I considered these data 'missing'. I investigated characteristics of individuals with BP value data using linkage between the primary care dataset and the UK Biobank baseline assessment dataset (see section 8.2.7, below, for these methods).

My next aim was to create a dataset of systolic blood pressure values. I chose to focus on systolic blood pressure, rather than diastolic blood pressure, because systolic BPV has been shown to associate more strongly with risk of stroke than diastolic BPV.(Rothwell et al. 2010b) VALUE1 was assumed to be systolic BP (SBP) and VALUE 2 was assumed to be diastolic BP (DBP) (Table 8.1, above). Having excluded all records with 'missing' VALUE1 values, I removed potential 'errors' in SBP data entry, as follows:

1) I excluded records where VALUE1 (presumed SBP) ≤ VALUE2 (presumed DBP).

2) If VALUE1 was present but VALUE2 was missing, I reviewed the Read code text definition, and included only those records which specified that the VALUE1 value was 'systolic BP'. There are twenty-two Read codes which specify (in their associated text definitions) that a value is 'systolic' or 'diastolic' BP. These Read codes are listed in the Appendix 8.4.2 (Table 8.15).

3) I investigated the range of remaining VALUE1 values, and excluded 'extreme' values (>300mmHg).

4) Finally, I grouped records by participant and event date to identify all VALUE1 values which were recorded at the same GP visit. Where multiple VALUE1 values

were entered at the same GP visit, I selected one at random, and excluded all others. This resulted in an SBP dataset with one coded SBP value per GP visit.

8.2.7 Feasibility of using primary care coded data to estimate visit-to-visit BPV

I aimed to explore the feasibility of using primary care coded data (in particular coded SBP values at successive GP visits) to estimate within-individual visit-to-visit SBP variability. In the long run, these data could be used to estimate BPV as an 'exposure' in UK Biobank and could be linked to outcomes, enabling investigation of potential associations between BPV and outcomes such a stroke.

I used a minimum of 3 GP visits with coded BP readings to estimate BPV. It has previously been shown that increasing the number of visits with BP readings (from 3 to 11) increases the reproducibility of BPV measurement.(Rothwell et al. 2010b) However, an optimum measurement period has not been assessed, and it has not yet been determined if there is a ceiling beyond which increasing the number of included BP measurements fails to increase the precision of visit-to-visit BPV estimation. I hypothesised that there may be an inverse association between the numbers of GP visits with coded BP readings, the time prior to recruitment within which these BP readings are recorded, and the numbers of individuals identified. I sought to determine the best balance between these factors. My ultimate aim was to maximise the number of individuals in whom visit-to-visit BPV could be measured prior to UKB recruitment, as this would increase future power to detect potential associations between BPV, risk of stroke, and its main types. It was important to keep both the numbers of BP readings used and the BP measurement periods (prior to UKB recruitment) the same for each individual, because both parameters may influence the strength of the association between BPV and risk of stroke (Chapter 7).

Number of individuals identified with 'x' coded BP values over 'y' time periods

I firstly aimed to investigate the number of participants I could identify with 'x' BP values over 'y' time periods prior to UK Biobank recruitment. I focused on estimating BPV in the period prior to UKB recruitment in order assess the potential of measuring BPV as an 'exposure'. In future studies of associations between BPV

and stroke outcomes, the period prior to UKB recruitment, per individual, would be defined as the 'exposure period', and the period after recruitment, per individual, would be defined as the 'follow-up' period.

I began by counting the number of GP visits per participant which included SBP values. I then restricted analyses to data collected prior to UK Biobank recruitment (GP visit date ≤ participant's baseline assessment date). Having explored the range of data available, I counted the numbers of participants who could be identified with (a minimum of) three, five, seven, nine, and eleven SBP readings (over the same number of successive GP visits). I then stratified these groups by the time prior to UK Biobank recruitment within which these readings were recorded (≤ 1 year, ≤ 3 years, ≤ 5 years, ≤ 10 years before the baseline assessment date).

Within individual frequency of coded SBP values in the coded primary care dataset.

I secondly aimed to explore the frequency of recorded SBP values in each participant's coded primary care record. In previous studies of the association between medium-to-long term BPV and stroke risk (prospective observational studies and clinical trials) the gaps between successive outpatient visits (and therefore successive BP readings) were standardised. I investigated within participant frequency of successive GP visits with coded BP values, as follows:

1) I calculated the gap (in days) between successive GP visits for each participant, using the coded event dates.

2) I calculated the median of the within participant gaps, per participant, prior to UK Biobank recruitment.

3) I grouped results into gaps of 6 month increments (ranging from ≥ 1 day but ≤ 6 months, to ≥ 6 years). There were no gaps of ≤ 1 day, because coded events with the same date and same participant identifier were assumed to have originated from the same GP visit (and I had already selected one adjudicated BP value from each GP visit at random, above).

4) I further stratified gaps of between ≥ 1 day and ≤ 18 months into groups of 30 day increments (from ≥1 and ≤30 days to ≥516 to ≤560 days).

Numbers of participants identified with 3 to 11 GP visits at specified intervals, prior to UKB recruitment.

I explored the numbers of participants I would include (in estimation of BPV prior to UKB recruitment) if both the number of BP readings used, and frequency of BP readings, were fixed. I first identified five groups of participants who had ≥4, ≥5, ≥6, ≥7, and ≥8 GP visits with coded SBP values prior to UKB recruitment. I then specified four gap ranges (based on the frequencies of BP measurements in previous studies, chapter 8). The gap ranges were: 1 to 30 days, 1 to 5 months, 5 to 9 months and 10 to 14 months between GP visits with coded SBP values. I searched for participants in each of the five groups, above, who had consecutive gaps of these durations. In order to maximise the number of participants identified with consecutive gaps of each duration I ignored GP visits which took place before the specified time interval (eg. I ignored a GP visit within 2 weeks of the last visit if the required gap was 1-5 months, and there was a subsequent GP visits in 2 months' time). GP visits were therefore not necessarily consecutive. Having identified numbers of participants with 3,4,5,6,and 7 gaps (duration ranges specified above), between any 4,5,6,7,and 8 GP visits prior to UKB recruitment, I calculated the median within participant gap for each group.

8.2.8 Characteristics of selected participants

Generalisability to the UK Biobank Welsh population

I investigated the characteristics of individuals identified with at least one coded SBP value in the UK Biobank Welsh primary care dataset, as well as characteristics of individuals with any coded data in the UK Biobank Welsh primary care dataset. I linked the primary care dataset to the UK Biobank baseline assessment dataset (individual records were linked by participant identifier). I extracted the following participant characteristics from the baseline assessment dataset (see Appendix 8.4.2, Table 8.16, for the codes selected): gender (male/female), mean recruitment age (years), Townsend deprivation score (range -6 to 12), self-reported smoking status (never smoked, current smoker, ex-smoker), and self-reported comorbidities

(including stroke, myocardial infarction, hypertension, diabetes, and renal disease). Self-reported comorbidities were based on answers given during a brief nurse led interview, which verified responses to the UKB baseline assessment touchscreen questionnaire.

I then investigated characteristics of individuals in six mutually exclusive groups: those with one coded SBP value, 2 values, 3 to 11 values, 12 to 20 values, 21 to 29 values, or >30 values, recorded at any time (before or after UKB recruitment).

Finally, I investigated characteristics of individuals with 'x' numbers of SBP readings in 'y' time periods prior to UKB assessment (numbers of readings and time periods were chosen to maximise the number of individuals included). To ensure that the same numbers of SBP readings were used for each individual in each time period (and to maximise the number of individuals who were able to provide SBP data for estimation of BPV), I selected the required number of SBP readings at random from all participants who had at least that number of readings in the specified time period. For example, to select individuals with 5 SBP readings within 5 years of UKB recruitment, I selected 5 SBP readings at random from all individuals identified with ≥5 SBP readings recorded 5 years before recruitment. I calculated the median gap between selected SBP readings for each participant in each group (with 'x' SBP readings over 'y' time periods), and then calculated the median within participant gap for each group.

Calculation of BP variability (BPV) and the variability in BPV amongst individuals with 'x' BP readings over 'y' time periods.

I calculated the within individual BPV (within individual SD in SBP), in each group with 'x' BP readings over 'y' time periods. I used this to infer the potential range of SBP values about the individual mean SBP. If there is sufficient variability in within individual BPV amongst included individuals it should still be possible to detect a potential association between increasing BPV and increasing risk of stroke, even if these individuals have higher BPV than the population as a whole.

As above, SBP values were selected at random to ensure the same number of SBP readings were used for each individual. I then calculated the mean and standard

deviation of the within individual BPV (mean BPV and SD BPV) for each group with ‘x’ BP readings over ‘y’ time periods. Analyses were performed in *Stata* (version 12).

Finally, I used the SD BPV to infer the likely range of BPV values in each group with ‘x’ BP readings over ‘y’ time periods. 95% of the range in BPV is the mean BPV for that group +/- 1.96 x SD BPV for that group:

95% BPV range = mean SBP +/- 1.96 x SD BPV

8.2.9 Agreement between UKB SBP measurements and primary care coded data

I used the intraclass correlation coefficient (ICC) to assess within individual agreement between systolic BP recorded at the UK Biobank baseline assessment and systolic BP recorded in primary care coded data. Two blood pressure readings were recorded during the UK Biobank baseline assessment. Measurement followed a standard protocol, which was published online.(http://www.ukbiobank.ac.uk/) For each participant I used linkages to UKB baseline assessment data to compare the mean SBP recorded at the baseline assessment to the mean of all SBP values recorded in the primary care coded data before or after this date (within 10 days, 30 days, 3 months, 6 months, 1 year, 18 months and 2 years). ICC and 95% confidence intervals were calculated in *stata* (version 12) using a 2-way random effects model.

8.3 Results

8.3.1 Describing the dataset

Amongst 9,947 participants in UK Biobank Welsh primary care coded dataset, there were ~ 5,000,000 coded events, with up to ~6,000 coded events per individual.

The number of coded events increased every year from 1980 to end 2012 (Figure 8.1).

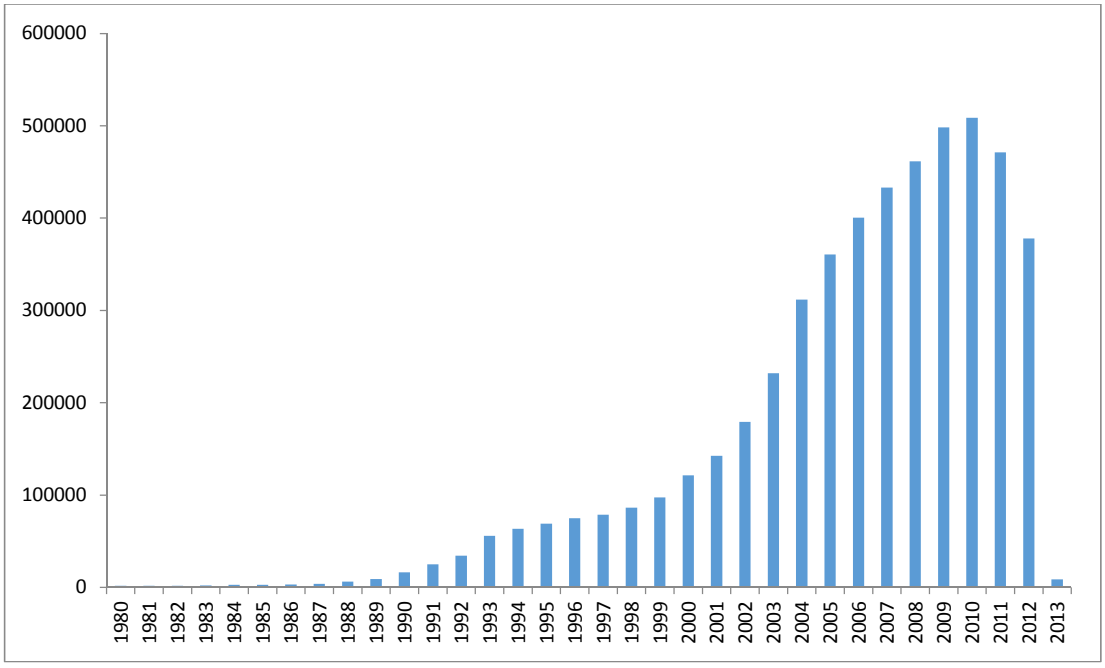

Figure 8.1 Number of coded events per year

y-axis: number of coded events

x-axis: year (from 1980 onwards).

The majority of participants (~80%) had coded primary care records which spanned a period of greater than 20 years (based on last event date – first event date)(Table 8.2). Few participants (~4%) provided coded data over shorter durations (≤ 5 years), and of these, only a minority (77 participants, 0.8%) had only one GP visit in the dataset.

Table 8.2 Time from first to last GP visit (per participant)

Time* (years)	Participants (n)	% Total (9,947)
≤1†	119†	1.2
1 to 5	289	2.9
5 to 10	420	4.2
10 to 20	1,204	12.1
>20	7,915	79.6

*Based on last event date – first event date.

†77 participants had only one event date recorded in the dataset (only one GP visit).

8.3.2 Characteristics of the denominator population

Characteristics of the 9,947 UKB Welsh participants with UKB linked coded primary care data are displayed in Table 8.3.

Table 8.3 Characteristics of the Welsh primary care population

Selected Characteristics*	UKB Welsh participants (n= 9,947)
Gender	45.5% Male
Recruitment age (mean, range)	56.1 (40 to 70)
Townsend score (mean, range)	-1.65 (-6.2 to 8.6)
Smoking status†	(%)
Never smoked	54.8
Current smoker	10.7
Ex-smoker	33.6
Co-morbidities‡	(%)
Stroke§	0.9
MI	0.6
Hypertension	24.5
Renal disease¶	0.1
Diabetes**	4.4

*Data from the UKB baseline assessment (see Appendix 8.4.2, Table 8.16 for the data fields used).

†Based on data from the baseline assessment touchscreen questionnaire (participant self-report). 0.4% missing or 'prefer not to answer'.

‡Data from participant self-report verified by nurse led interview.

§Ischaemic stroke, ICH, SAH, or unspecified stroke.

¶Renal/kidney failure (on dialysis or not on dialysis).

**Type 1 or type 2 diabetes.

Participants were 46% male, and aged between 40 and 70 years (mean age 56). The Townsend score for deprivation was -1.65 (range -6.2 to 8.6), with a lower score representing less deprivation. ~55% of participants were lifelong non-smokers, (based on self-report in the UKB baseline assessment touchscreen questionnaire), and the prevalence of self-reported comorbidities (verified during the UKB nurse led verbal interview) were ~25% hypertension, ~2.5% stroke/MI, 0.1% renal failure, 4.4% diabetes.

8.3.3 Identification of blood-pressure related codes

I identified ~ 296 Read codes related to blood pressure. BP related Read codes (excluding codes for drug prescriptions) accounted for ~ 3% of all coded events in the UKB Welsh coded primary care dataset (n=171,142 coded events out of a total ~5,000,000). Blood pressure value codes accounted for the majority of all BP related coded events (87%), followed by blood pressure monitoring codes (9%), hypertension diagnosis codes (3%), and hypertension administration codes (0.65%). Hypotension diagnosis codes accounted for only 0.05% of BP-related coded events (see methods 8.2.5 for a description of these Read code categories).

Table 8.4 shows the most frequently used BP Read codes in each of the four main categories. Only seven codes together accounted for 94% of all BP-related coded events (indicated by shading in Table 8.4). The frequencies of every BP related Read code (most of which accounted for <1% of all BP related coded events in the dataset) are listed in the Appendix 8.4.2 (Table 8.17).

Table 8.4 Most frequently used BP Read codes

Code	Code definition	Code frequency*	% of category total†
Hypertension diagnosis codes			
G2…	Hypertensive disease	2,074	37.5
G20..	Essential hypertension	3,148	56.9
G20z.	Essential hypertension NOS	155	2.8
G201.	Benign essential hypertension	76	1.4
Hypertension administration codes‡			
9N1y2.	Seen in hypertension clinic	291	26.1
67H8.	Lifestyle advice regarding hypertension	285	25.6
9N03.	Seen in hypertension clinic	256	22.9
14A2.	History of hypertension	180	16.1
Blood pressure monitoring codes			
662..	Hypertension monitoring	5,739	39.1
662P.	Hypertension monitoring	2,659	18.1
662d.	Hypertension annual review	1,078	7.3
662V.	Blood pressure monitoring	983	6.7
9O14.	Hypertension monitoring-1st letter	728	4.9
Blood pressure value codes			
246..	O/E blood pressure reading	144,868	96.7
2464.	O/E blood pressure normal	1,790	1.2

Shading indicates Read codes which together accounted for ~ 94% of all BP related coded events (excluding codes for medication prescriptions).

*Amongst 171,142 BP related codes.

†% of total number of codes in each category.

‡This category excludes hypertension monitoring codes.

8.3.4 Individuals with probable or possible hypertension

9,439 participants (95% of 9,947 UKB Welsh participants) had at least one BP related Read code. Of these participants, 29% had at least one Read code indicating a 'probable' or 'possible' hypertension diagnosis (n=2,713 out of 9,439, Figure 8.2); 24% had at least one 'hypertension diagnosis' code' indicating probable hypertension (n=2,237), and the remaining 5% (n=476) had at least one 'hypertension, other' code (administration codes or BP monitoring codes indicating 'possible' hypertension).

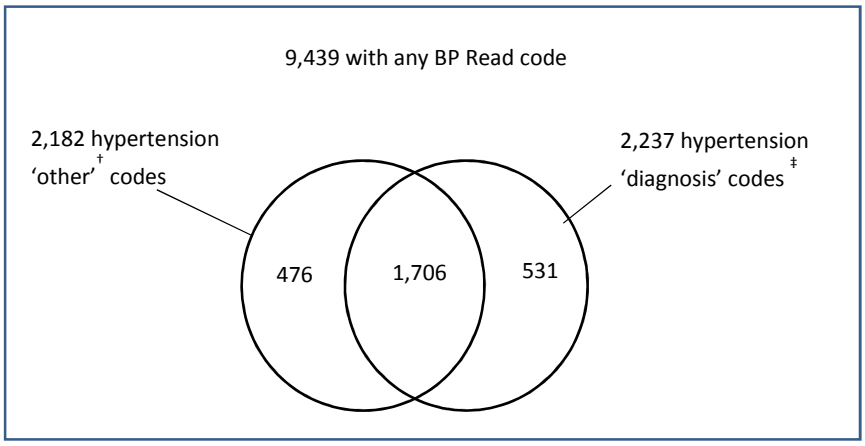

Figure 8.2 Proportion of 9,439 individuals with BP related Read codes who had 'probable' or 'possible' hypertension*

*Does not include analysis of BP values.

†Codes for 'possible hypertension', which include 'administration codes' and selected 'blood pressure monitoring codes' (see Appendix 8.4.2, Table 8.14, for codes selected).

‡Codes for 'probable hypertension', which include all 'hypertension diagnosis' codes (Appendix 8.4.2, Table 8.14).

The majority of participants had at least one BP value code in the coded primary care dataset. In total, 95% of Welsh UKB participants had ≥1 BP value code, which included 99.9% of individuals with 'probable hypertension' and ~93% of individuals without 'probable hypertension'.(Figure 8.3). Around 27% of Welsh UKB participants had at least one BP monitoring code, and, 99.9% of these had at least one BP value code.

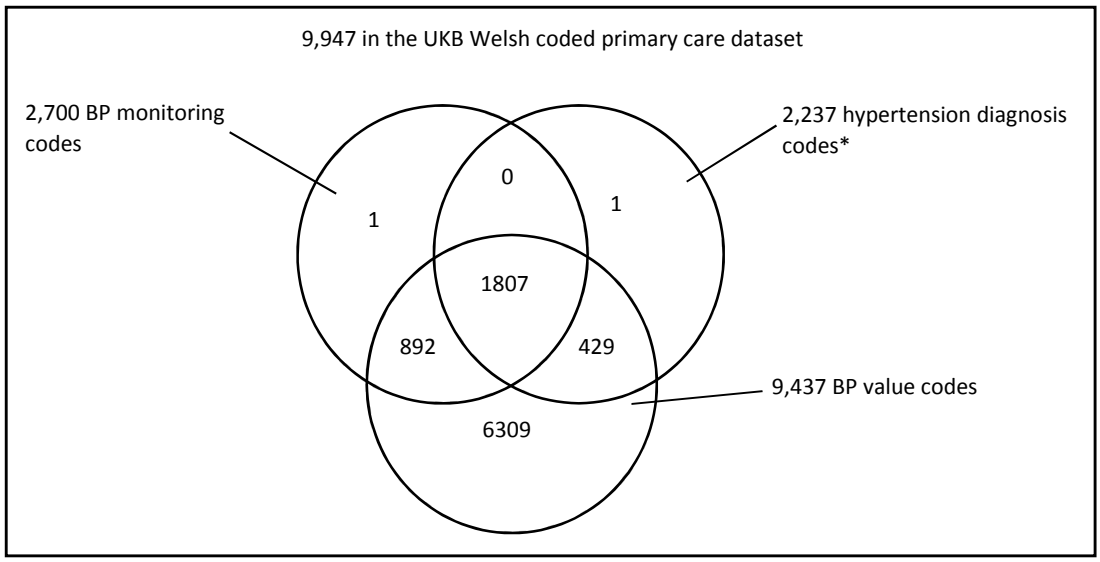

Figure 8.3 Proportions of individuals (n=9,947) in the UKB Welsh coded primary care dataset who have hypertension diagnosis codes, BP monitoring codes, and/or BP value codes.

*Codes for 'probable hypertension'.

8.3.5 Creating an SBP value dataset

Missing data

SBP values were missing in 14% of coded events (n=23,051 out of 164,424 coded events) with BP monitoring or BP value Read codes. The majority of 'missing data' came from events coded with BP monitoring Read codes; only 2.9% of these (n= 452 out of 15,483) had associated SBP values. Around 95% of events coded with BP value Read codes had SBP values (n=140,921 out of 148,941 records). There were no SBP values associated with 'hypertension diagnosis' or 'hypertension administration' Read codes, but, as illustrated in Figure 8.3, 99.9% of individuals with 'hypertension diagnosis' and, or 'hypertension administration' Read codes also had BP value Read codes.

Adjudication of SBP values

Figure 8.4 displays the process of adjudication of the 141,374 records with potential SBP values. In 99% of cases (n= 140,066), both VALUE1 (presumed SBP) and VALUE2 (presumed DBP) were present. I excluded 60 records where VALUE1 was ≤ VALUE2. I retained ~ 27% of records which only included VALUE1 values (n=

352 out of 1,308) because 'systolic BP' was specified in the Read code text definition (see Appendix 8.4.2, Table 8.15, for these Read codes).

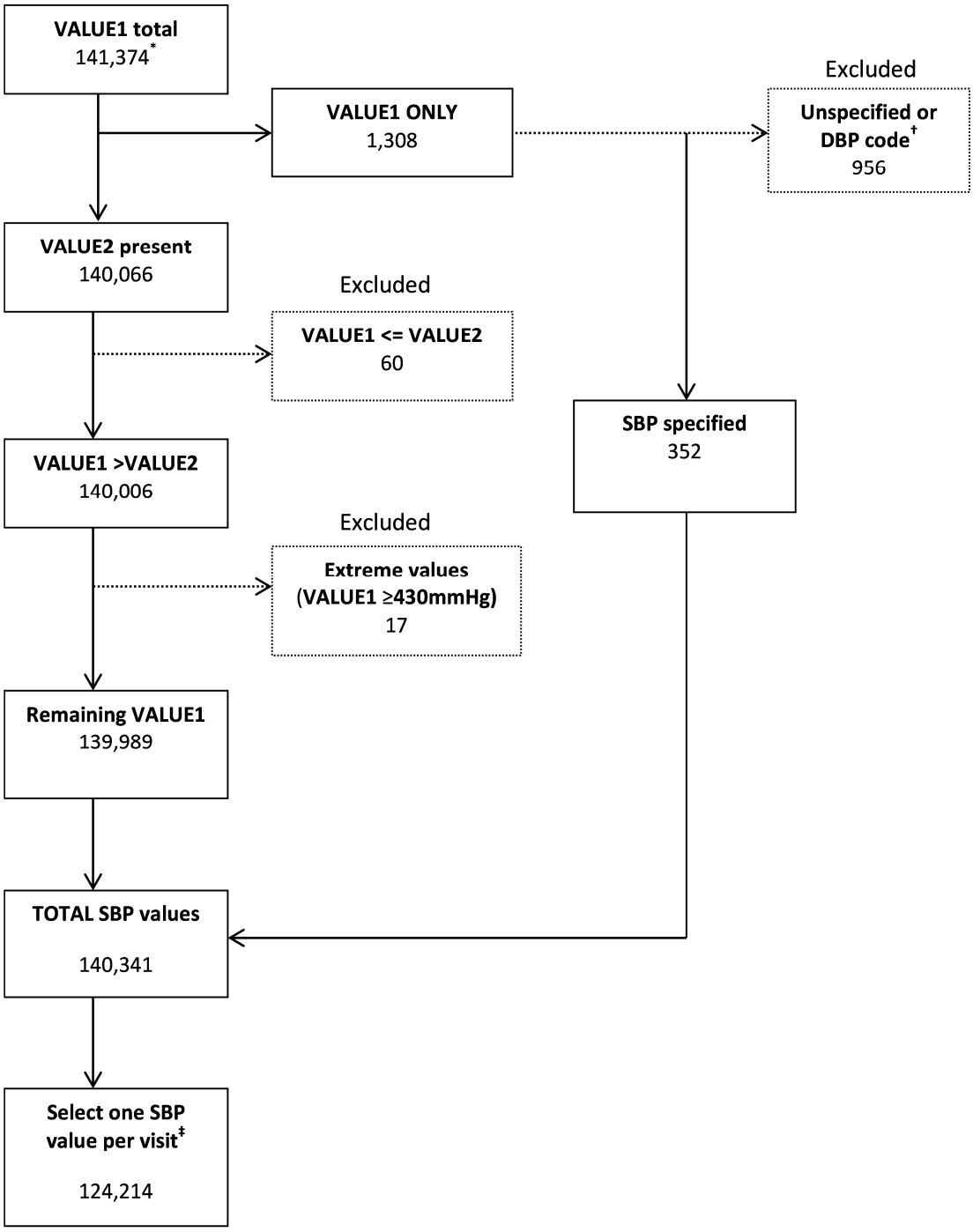

Figure 8.4 Adjudication of potential systolic BP (SBP) values

*Total non-missing records with VALUE1 values (potential SBP).

†Includes VALUE1 values specified as diastolic BP readings (DBP)- see Appendix 8.4.2, Table 8.15 for these Read codes- or includes unspecified VALUE1 values (ie. readings not specified as systolic or diastolic BP).

‡Selected one value at random from each GP visit, per participant.

The range and frequency of the remaining 140,006 VALUE1 values are presented in Figure 8.5. The majority of these potential SBP values (99.9%) ranged from 60 to 260mmHg. There were only 27 records with SBP values outside of this range, and from these, I excluded 17 records with 'impossible' values (range 430 to 14130mmHg). I retained 10 records with SBP values ranging from 10 to 58mmHg. There was marked 'digit preference' in the dataset, such that SBP values were most frequently recorded to the nearest 10mmHg (eg. 110mmHG, 120mmHg, 130mmHg etc.).

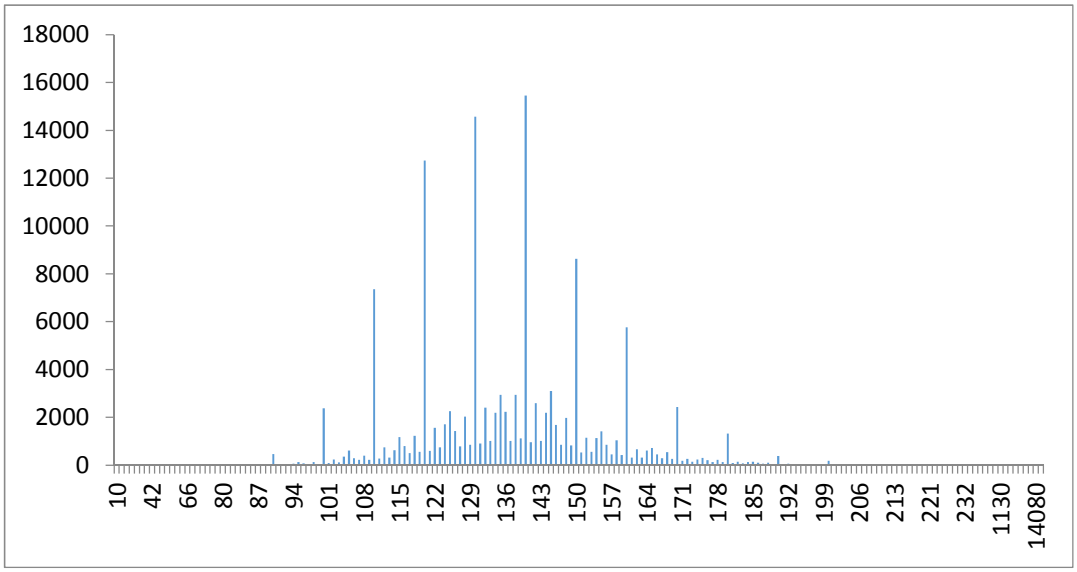

Figure 8.5 Range of SBP values*

y-axis: Frequency (out of 140,006 VALUE1 values in the dataset)

x-axis: Systolic BP (mmHg)

Selecting one SBP value per GP visit

When data were grouped by participant identifier and event date, there were 124,214 separate GP visits. The number of SBP values entered per GP visit ranged from 1 to 19. In the final step of adjudication I selected one SBP value at random from each

GP visit. Around 90% of GP visits only recorded one SBP value. Amongst GP visits with multiple SBP values (≥2 SBP values) 76% had the same value recorded multiple times.

8.3.6 Feasibility of estimating BPV using primary care coded data

Characteristics of the denominator population

In the end there were 9,288 UKB Welsh participants with at least one systolic BP reading in the final adjudicated coded primary care dataset. I had excluded 149 participants (1.6% of all 9,437 participants with BP value Read codes) because of missing VALUE1 data and/or errors in VALUE1 data entry (see Figure 8.4). Characteristics of the 9,288 individuals with at least one adjudicated SBP value in the UKB Welsh primary care coded dataset are displayed in Table 8.5. Characteristics of 9,947 individuals with any data (at least one coded event) in the UKB Welsh population are displayed for comparison.

Table 8.5 Characteristics of UKB Welsh participants with ≥1 coded (and adjudicated) SBP value

Characteristics[*]	Welsh participants with ≥1 coded SBP value (n= 9,288)	Welsh participants with continuous UKB registration (n= 9,947)
Gender	45.5% Male	45.5% Male
Mean recruitment age	56.1 (*range 40 to 70*)	56.1 (*range 40 to 70)*
Townsend score	*-1.69 (range -6.2 to 8.6)*	-1.69 (*range -6.2 to 8.6)*
Smoking status[†]	*(%)*	(%)
Never smoked	54.8	54.8
Current smoker	10.7	10.7
Ex-smoker	33.9	33.9
Co-morbidities[‡]	*(%)*	*(%)*
Stroke[§]	0.9	0.9
MI	0.6	0.6
Hypertension	25.1	24.5
Renal disease[¶]	0.1	0.1
Diabetes[**]	4.5	4.4

[*]Data from the UKB baseline assessment (see Appendix 8.4.2, Table 8.16 for the data fields used).

[†]Based on data from the baseline assessment touchscreen questionnaire (participant self-report). 0.4% missing or 'prefer not to answer'.

[‡]Data from participant self-report verified by nurse led interview.

[§]Ischaemic stroke, ICH, SAH, or unspecified stroke.

[¶]Renal/kidney failure (on dialysis or not on dialysis).

**Type 1 or type 2 diabetes.

There was no material difference when characteristics of participants with ≥1 coded SBP value were compared to characteristics of all participants with any coded primary care record in the UKB Welsh population (the former population being a large sub-set of the latter).

Characteristics of participants with at least one coded SBP value, stratified by the number of successive GP visits with coded SBP values

The number of successive GP visits with coded SBP values ranged from 1 to 242 per individual (median 10, IQR 4 to 21). The numbers of individuals with 1 to 20 GP visits with coded SBP values, and with ≥2 to ≥20 SBP values (out of all 9,288 with ≥1 SBP value) are displayed in Table 8.6. 83% of participants had ≥3 SBP values (sufficient to estimate BPV).

Table 8.6 Proportions of participants (out of 9,288 with ≥1 coded SBP value in the Welsh primary care dataset) with 1 to 20, and ≥1 to ≥20 coded SBP values

Number of SBP readings	Participants (n)	(%)	Range of SBP readings	Participants (n)	(%)
1	902	(9.8)	≥1	9,288	(100)
2	717	(7.7)	≥2	8,368	(90)
3	660	(7.1)	≥3	7,669	(83)
4	648	(6.9)	≥4	7,009	(75)
5	543	(5.8)	≥5	6,361	(68)
6	473	(5.1)	≥6	5,818	(63)
7	439	(4.7)	≥7	5,345	(58)
8	417	(4.5)	≥8	4,906	(53)
9	417	(4.5)	≥9	4,489	(48)
10	300	(3.2)	≥10	4,131	(44)
11	258	(2.8)	≥11	3,831	(41)
12	259	(2.8)	≥12	3,573	(38)
13	221	(2.4)	≥13	3,314	(36)
14	175	(1.9)	≥14	3,093	(33)
15	167	(1.8)	≥15	2,918	(31)
16	163	(1.8)	≥16	2,751	(30)
17	147	(1.6)	≥17	2,588	(28)
18	177	(1.9)	≥18	2,441	(26)
19	136	(1.5)	≥19	2,264	(24)
20	107	(1.2)	≥20	2,128	(23)

Characteristics of individuals, stratified by the number of coded SBP values present in the primary care dataset.

Individuals with increasing numbers of coded SBP values (from ≥12 to ≥21 to ≥30) appeared to have a greater prevalence of hypertension and diabetes than individuals with fewer numbers of coded SBP values (<12 SBP values). Table 8.7 displays characteristics of the population with coded SBP values (n=9,288), stratified by the total number of SBP values per individual. These values were recorded at any time (before or after UKB recruitment) and at any frequency (any time periods between successive BP values). Participants with from 3 to 11 visits with BP values in their coded primary care record did not differ significantly in terms of cardiovascular risk factors from participants with only one or two visits with BP values in their coded primary care record. In fact, participants with ≤11 BP values in their primary care record had a lower prevalence of vascular risk factors compared to the denominator population (24.5% hypertension, 4.4% diabetes, 0.1% renal disease and 1.5% cardiovascular disease, see Table 8.3). Only those participants with ≥21 visits with BP values in their coded primary care record appeared to have increased prevalence of vascular risk factors compared to the denominator population.

Table 8.7 Characteristics of participants with 1 to >30 SBP values in the coded primary care dataset

SBP values* (n)	Participants† (n)	Hypertensive (% & 95% CI)		Diabetic (% & 95% CI)		Renal failure (% & 95% CI)		CVD‡ (% & 95% CI)	
1	902	12.4	(10.4 to 14.7)	1.8	(1.1 to 2.9)	0	(0 to 0)	0.7	(0.3 to 1.4)
2	717	9.6	(7.6 to 12.0)	1.1	(0.5 to 2.1)	0.1	(0.02 to 0.8)	0.7	(0.3 to 1.6)
3-11	4,096	9.8	(8.9 to 10.7)	1.3	(0.9 to 1.7)	0.05	(0.01 to 0.2)	1.0	(0.7 to 1.4)
12-20	1,552	29.1	(26.9 to 31.4)	4.9	(2.8 to 4.7)	0.06	(0.01 to 0.4)	2.6	(1.9 to 3.5)
21-29	913	56.3	(53.1 to 59.5)	9.2	(7.5 to 11.3)	0.2	(0.06 to 0.8)	2.4	(1.6 to 3.6)
≥30	1,108	70.7	(42.3 to 48.2)	16.1	(14.0 to 18.3)	0.4	(0.1 to 0.9)	2.3	(1.5 to 3.3)

*Number of SBP values in the dataset. Each SBP value came from a separate GP visit.

†Out of 9,288 with at least one coded SBP value.

‡Cardiovascular disease: myocardial infarction and stroke.

Numbers of participants with ≥ 3 GP visits with coded SBP values, stratified by time prior to UK Biobank recruitment

There were 8,827 participants (~89% of 9,947 in the Welsh UKB population) who had at least one coded SBP value recorded prior to UKB recruitment. Table 8.8 shows the numbers of participants who had at least 3 coded SBP values (ranging from ≥3 to ≥11) stratified by the time prior to UKB recruitment.

The numbers of participants identified with 'x' BP readings over 'y' time periods prior to UKB recruitment decreased as both the number of SBP values required was increased and the measurement period was decreased. Around 68% of the Welsh UKB population (n= 9,947 who provided linked coded primary care data) had at least 3 coded SBP values in the dataset any time before UKB recruitment (Table 8.8). However, this decreased to ~52% with ≥5 SBP values, ~40% with ≥7 SBP values, 32% with ≥9 SBP values, and ~27% with ≥11 SBP values any time before recruitment.

Few participants in each of these categories had all of their SBP values recorded within 1 to 2 years prior to UKB recruitment (range 1.3 to 39% of participants in each category with ≥3 to ≥ 11 SBP values). Five years prior to UKB recruitment appeared to be the optimum exposure period to maximise the number of participants included, without using too long a period of time (ie. extending the SBP measurement period up to 10 years prior to UKB recruitment, or using an 'unspecified' time prior to UKB recruitment). This time period identified at least 50% of participants in each category with ≥3 to ≥11 SBP values (See Table 8.8, shaded rows). The proportion of the total UKB Welsh population (n=9,947 with coded primary care data) who had ≥3 to ≥11 coded SBP values recorded in the dataset 5 years before recruitment ranged from 14 to 47%.(Table 8.8).

Table 8.8 Numbers of participants with ≥3 to ≥ 11 SBP values stratified by time before UKB recruitment during which these values were recorded

SBP values* (range)	Time prior to recruitment (years)†	Participants (n)	% total in each category	% 9,947 in UKB‡
≥3	≤1	1,170	17.3	11.8
≥3	≤2	2,616	38.6	26.3
≥3	≤3	3,490	51.5	35.1
≥3	≤5	4,658	68.8	46.8
≥3	≤10	5,836	86.1	58.7
≥3	Unspecified	6,775 (total)	100	68.1
≥5	≤ 1	429	2.3	4.3
≥5	≤ 2	1,393	26.9	14.0
≥5	≤ 3	2,224	42.9	22.4
≥5	≤ 5	3,163	61.1	31.8
≥5	≤10	4,197	81.0	42.2
≥5	Unspecified	5,179 (total)	100	52.1
≥7	≤ 1	176	4.4	1.8
≥7	≤ 2	738	18.4	7.4
≥7	≤ 3	1,440	35.9	14.5
≥7	≤ 5	2,352	58.6	23.6
≥7	≤10	3,231	80.6	32.5
≥7	Unspecified	4,011 (total	100	40.3
≥9	≤ 1	78	2.8	0.8
≥9	≤ 2	424	13.1	4.3
≥9	≤ 3	917	28.2	9.2
≥9	≤ 5	1,845	56.8	18.5
≥9	≤10	2,550	78.5	25.6
≥9	Unspecified	3,248 (total)	100	32.6
≥11	≤ 1	35	1.3	0.4
≥11	≤ 2	251	9.3	2.5
≥11	≤ 3	606	22.4	6.1
≥11	≤ 5	1,421	52.5	14.3
≥11	≤10	2,122	78.3	21.3
≥11	Unspecified	2,709 (total)	100	27.2

*Each coded SBP value was recorded at a separate GP visit, therefore to have ≥3 SBP values, an individual would have to have had ≥3 GP visits where BP was measured and recorded in the coded primary care dataset.

†Time periods prior to UKB recruitment within which SBP values are recorded.

‡Proportion of 9,947 participants in the UKB Welsh primary care coded dataset

Frequency of successive GP visits with coded SBP values

Around 88% of UKB participants with coded SBP values prior to UKB recruitment (n=7,748 out of 8,827) had at least two successive GP visits with coded SBP values. Amongst these individuals, the median time, per participant, between successive GP visits with coded SBP values ranged from 1 to 8,893 days (median 326 days, IQR 120 to 808 days). The most common gap (median time, per participant) between GP visits was from 1 day to 6 months long (representing ~36% of all gaps, Table 8.9), and the majority of gaps were less than eighteen months in duration (~65% of all gaps, Table 8.9)

Table 8.9 Range of time (median years per participant) between successive GP visits with coded SBP values

Gap* (years)	Gap frequency	% all gaps
≥1 day to ≤ 6 months	2,765	35.7
>6 months to ≤ 1	1,405	18.1
>1 to ≤ 1.5	880	11.4
>1.5 to ≤ 2	571	7.4
>2 to ≤ 2.5	399	5.1
>2.5 to ≤ 3	306	3.9
>3 to ≤ 3.5	285	3.7
>3.5 to ≤ 4	140	1.8
>4 to ≤ 4.5	142	1.8
>4.5 to ≤ 5	128	1.7
>5 to ≤ 5.5	100	1.3
>5.5 to ≤ 6	110	1.4
>6	517	6.7
Total	7,748†	100

*Median time (per participant) between successive GP visits with coded SBP values.

†Total number of participants with at least 2 successive GP visits with coded SBP values prior to UKB recruitment.

Attempting to specify the median within participant gap between GP visits with coded SBP values identified relatively small numbers of participants. Only 5% of all within participant median gaps were ≤ 1 month long (between 1 to 30 days), 9%

were around 4 months (>90 to ≤ 150 days), 10% were around 6 months (>150 to ≤210 days) and 10% were around 1 year (>300 to ≤425 days). The frequency of gaps (from ≥1 day to ≤ 18 months, in 30 day increments), and numbers of participants with these gaps (out of 5,050 total) is presented in the Appendix 8.4.2 (Table 8.18).

Table 8.10 presents numbers of participants who had consistent gaps between any 4 to any 8 of their GP visits with coded SBP values prior to UKB recruitment (amongst those who had ≥4 to ≥8 GP visits), stratified by the gap duration.
Out of 5,924 participants who had ≥4 GP visits with coded SBP values prior to UKB recruitment, up to 40% (range 12 to 39%) had consistent gaps between any four of their GP visits, and this represented only 7 to 23% of UKB Welsh participants with coded primary care data. (Table 8.10) Even fewer participants had larger numbers of GP visits (from ≥5 to ≥8 with coded SBP values prior to UKB recruitment), and fewer proportions of these participants had consistent gaps between any 5 to 8 of their GP visits.

Furthermore, the longer the gap duration specified, the fewer the number of participants identified with consistent gaps of that duration (Table 8.10). Within each category (4 to 8 GP visits with consistent gaps), the numbers of participants identified fell by about 89% on average (range 67 to 99%) when the specified gap duration was increased from ~2 weeks (range 1-30 days) to ~1 year (range 10-14 months). Specifying the gap between GP visits therefore resulted in exclusion of a large proportion of UKB participants, who would otherwise have contributed 4 to 8 SBP values for estimation of visit-to-visit BPV prior to UKB recruitment.(Table 8.10).

Table 8.10 Number of participants who had consistent gaps between GP visits[*] prior to UKB recruitment, stratified by the number of GP visits and the specified gap duration.

Gap[†] (range & median)	4 GP visits (n & % participants)[‡]	5 GP visits (n & % participants)[‡]	6 GP visits (n & % participants)[‡]	7 GP visits (n & % participants)[‡]	8 GP visits (n & % participants)[‡]
1-30 days (14 days)	2,110 (21)	1,395 (14)	880 (9)	546 (6)	335 (3)
1-5 months (77 days)	2,281 (23)	1,244 (13)	702 (7)	397 (4)	233 (2)
5-9 months (192 days)	1,567 (16)	568 (6)	261 (3)	127 (1)	64 (0.6)
10-14 months (357 days)	683 (7)	151 (2)	28 (0.3)	14 (0.1)	3 (0.03)
Unspecified[§]	≥4 GP visits 5,924 (60)	≥5 GP visits 5,179 (52)	≥6 GP visits 4,548 (46)	≥7 GP visits 4,011 (40)	≥8 GP visits 3,623 (36)

[*]GP visits were not successive. I excluded interim GP visits (eg. if gaps of 1-5 months were required between visits, I ignored visits with coded SBP values <1 month after the previous visit with coded SBP values).

[†]Specified time between 'n' GP visits with coded SBP values

[‡]% of 9,947 in the Welsh coded primary care dataset.

[§]Unspecified gap duration. These are the denominator populations: the total numbers of participants with ≥4 to ≥8 GP visits with coded SBP values prior to UKB recruitment. In order to identify the maximum number of participants with consistent gaps between 4 to 8 GP visits, I had to include those with ≥4 to ≥8 GP visits, and then exclude interim visits (described above).

Characteristics of participants with 3 to 11 SBP readings, 5 years before UKB recruitment

Table 8.11 displays characteristics of selected participants who had ≥3 to ≥11 coded SBP values recorded within the 'optimum time period': 5 years before UKB recruitment.

Table 8.11 Characteristics of participants with ≥3 to ≥11 SBP values within 5 years of UKB recruitment, compared to the denominator population.

SBP values[*] (range)	Participants[†] (n & %)	% Male	% HTN[‡]	% DM[‡]	% CVD[‡]	Selected SBP[§] (n)	Median gap[¶] (days)
-	9,947 (100)	45.5	24.5	4.4	1.5	-	-
≥3	4,658 (47)	43.4	39.8	7.5	2.2	3	413
≥5	3,163 (32)	45.1	53.1	10.4	2.6	5	224
≥7	2,352 (24)	46.5	63.4	12.9	2.8	7	163
≥9	1,845 (19)	47.6	68.6	14.5	2.7	9	122
≥11	1,421 (14)	47.9	72.2	15.1	2.5	11	98

[*]Range of SBP values, per participant, available 5 years prior to UKB recruitment

[†]Number of participants (% out of 9,947 with any coded data in the UKB Welsh primary care dataset) with the specified range of SBP values. Participants included in each row are subsets of participants in the rows above.

[‡]HTN= hypertension, DM= Diabetes (type 1 or type 2), CVD= cardiovascular disease, which included myocardial infarction or stroke. Results based on participant self-report, confirmed at the UKB baseline assessment nurse led interview.

[§]SBP values were selected at random from each participant, generating a group of participants with the same number of SBP values (from 3 to 11) within 5 years of UKB recruitment.

[¶]Median of the within participant median gaps (within participant median gap is the median of all gaps, per participant, between 'n' selected SBP values)

Using coded primary care data ≤5 years prior to recruitment, it was possible to estimate within participant BPV in up to 47% of UKB Welsh participants (n=4,658 out of 9,947). As the number of GP visits with coded SBP values was increased from ≥3 to ≥11, the proportion of participants with sufficient data to estimate BPV fell from 47 to 14%. However, absolute numbers of participants were high, and >1,000 to >2,000 still had sufficient data to estimate BPV using SBP values from 7 to 11 separate GP visits 5 years before recruitment. The median gap between GP visits, per participant, fell as the number of GP visits with coded SBP values was increased. Although one might argue that standardising the number of SBP values used, per participant, meant loosing data (eg. selecting 3 SBP values at random from all participants who had ≥3 SBP values 5 years before recruitment, and dropping any

additional SBP values), it was important to do this to avoid bias. It is known that the precision of BPV estimation increases with the number of SBP values used, and the strength of the observed association between BPV and stroke increases with the number of SBP values used to calculate it.(Rothwell et al. 2010b) It is also quite possible that participants with increased BP variability have an increased number and frequency of BP values in the dataset (compared to those with reduced BP variability). If the number of readings used was not standardised between individuals, those with more BP values (and hence those with potentially higher BPV), would demonstrate a stronger association between BPV and stroke, and this might falsely increase the observed association between increasing BPV and increasing stroke risk. Although standardising the number of BP values used to calculate BPV per individual does not change the fact those with higher number of values to begin with are likely to have an increased prevalence of vascular risk factors (eg. hypertension, diabetes etc.), these factors can be adjusted for more easily in subsequent analyses of the associations between increasing BPV and risk of stroke.

Generalisability of populations with 3 to 11 coded SBP values, 5 years before UKB recruitment

Increasing the number of SBP values required to estimate BPV reduced the generalisability of the populations selected. The proportion of participants with hypertension increased from ~40% to ~72%, diabetes from ~8% to 15%, and cardiovascular disease (MI or stroke) from 2.2 to 2.5% as the number of SBP values used to estimate BPV was increased from 3 to 11 within 5 years of UKB recruitment (Table 8.11). Overall, the participants in these selected groups had higher proportions of comorbidities than the original UKB Welsh population (n=9,947 which comprised 25% hypertension, 4.5% diabetes, and 1.5% CVD).

Mean and SD in blood pressure variability (BPV) amongst participants with 3 to 11 coded SBP values, 5 years before UKB recruitment

Within participant blood pressure variability (BPV) appeared to increase as the number of SBP values used to estimate BPV increased. When the number of SBP values used to estimate SD in SBP increased from 3 to 11, the mean BPV (the mean

of within participant standard deviations in SBP) increased from 10.7mmHg to 13.6mmHg. (Table 8.12). This could be due to increased precision of BPV estimation or due to selection of individuals with higher mean BP (and BP variability is known to increase as mean BP increases). The variability in BPV between selected participants decreased (SD in within participant BPV fell from 6.8mmHg to 4.8mmHg). Assuming that BPV is normally distributed, then 95% of all BPV values should fall within 1.96 standard deviations of the mean BPV in each group of participants. Therefore, amongst participants selected with 11 SBP values 5 years before recruitment, 95% of the range in BPV should be 13.6mmHg +/- 1.96 x 4.8mmHg = from 4.2mmHg to 23.2mmHg, or, in other words, 95% of the within participant SD SBP should range from 4.2mmHg to 23.2mmHg. This should be sufficient variability in between participant BPV to be able to detect a potential association between BPV and risk of stroke.

Table 8.12 Mean and variability in BPV (SD BPV) amongst selected participants with 3 to 11 SBP values 5 years before UKB recruitment.

SBP values[*] (n)	Participants (n)	Mean BPV[§]	SD BPV[†] (mmHg)
3	4,658	10.68	6.8
5	3,163	12.0	5.7
7	2,352	12.8	5.3
9	1,845	13.2	4.9
11	1,421	13.6	4.8

[*]SBP values selected at random from participants with ≥3 to ≥11 SBP values within 5 years of UK Biobank recruitment.

[§]Mean within participant SD in SBP

[†]SD in within participant SD in SBP

8.3.7 Comparison between UKB baseline assessment BP and BP values in the coded primary care dataset.

Agreement between mean SBP using coded SBP values within 10 days to 2 years of the UK Biobank assessment, and the mean baseline assessment SBP was generally good; ICC ranged from 0.478, (95% CI 0.439 to 0.515) for mean coded SBP within

30 days of baseline assessment, to 0.534 (95% CI 0.517 to 0.551) for mean coded SBP within 2 years of the baseline assessment (Table 8.13).

Table 8.13 Within participant agreement between mean SBP in the coded primary care dataset[*], and mean SBP measured at the UKB baseline assessment, stratified by time from the baseline assessment within which mean coded SBP was measured.

Time[†]	Participants (n & %)[‡]	ICC (95% CI)
10 days	816 (9)	0.514 (0.461 to 0.562)
30 days	1,641 (18)	0.478 (0.439 to 0.515)
90 days	3,137 (34)	0.460 (0.433 to 0.487)
6 months	4,347 (47)	0.486 (0.463 to 0.509)
12 months	5,518 (59)	0.511 (0.492 to 0.530)
18 months	6,243 (67)	0.522 (0.504 to 0.540)
24 months	6,716 (72)	0.534 (0.517 to 0.551)

[*]Mean of 'n' values, per participant, which were recorded 10 days to 24 months before or after the UK Biobank baseline assessment. If only one SBP value was recorded in the coded primary care dataset in the specified time period, this single value was used.

[†]Time from UKB baseline assessment within which mean coded SBP was measured, per participant.

[‡]% of 9,288 participants who had at least one coded SBP value recorded before or after UKB recruitment.

8.4 Discussion

Almost 50% of UK Biobank Welsh participants with linked, coded, primary care data had sufficient GP visits with coded SBP values (at least 3 GP visits with coded values, up to 5 years before UK Biobank recruitment) to be able to estimate within individual visit-to-visit BPV. Around a third to a quarter of these participants had 5 to 7 GP visits with coded SBP values up to 5 years before UKB recruitment. Fewer participants were included if the gap between successive GP visits with coded SBP values was fixed. Participants in whom BPV could be measured had higher proportions of hypertension (range ~40% to ~72%), diabetes (range ~8% to ~15%), and MI/stroke (range ~2% to ~3%) than the original UKB Welsh population (total 9,947 with linked, coded primary care data). Selecting participants based on increasing numbers of GP visits with coded SBP values (from 3 to 11) increased both the prevalence of vascular risk factors and the mean BPV, but reduced the within participant SD BPV (SD in within participant BPV). However, there was still sufficient variation in BPV between individuals (SD BPV ranged from 5.3 to 6.8mmHg amongst participants with 3 to 11 BP readings, 5 years before UKB recruitment) to recommend using these selected populations in future studies of association between BPV and stroke. Furthermore, there was reasonable within participant agreement between mean BP recorded at the UK Biobank baseline assessment and mean BP estimated using coded primary care data (ICC for agreement with baseline assessment mean BP ranged from 0.46 to 0.53 when mean BP was estimated using coded primary care data within 2 weeks to 2 years of the assessment). This suggests that BP measurements extracted from coded primary care data are a reasonable approximation to standardised BP measurements. In conclusion, linkage to coded primary care data is an efficient method of estimating within individual visit-to-visit BP variability, particularly in large populations.

The numbers of participants in whom BPV could be measured in this study (~2,000 to up to ~5,000), and the measurement periods (up to 5 years) are similar to those used in previous studies of the associations between BPV and risk of stroke (clinical trials or observational studies, Chapter 6). Although it was not possible to specify the gap between successive GP visits with coded BP values (because too few participants were identified), the median of the median within participant gap (for

those with 3 to 7 successive BP readings within 5 years of recruitment) was ~5 months to ~1 year, which was also similar to the frequency of GP visits used in previous studies. This suggests that the coded data available in primary care datasets should be at least sufficient to replicate previous studies of associations between BPV and stroke in the UK Biobank population. However, UK Biobank has the additional advantages of more accurate and detailed stroke outcomes adjudication methods compared to many previous studies, and will offer the potential to adjust for potential confounding factors, like renal failure, which were not accounted for in the majority of previous analyses of the associations between BPV and risk of stroke.

Coded primary care data have a number of potential advantages compared to data that have been used previously. Firstly, the prevalence of hypertension, diabetes, and cardiovascular disease (MI or stroke) was lower in populations selected using coded primary care data than in previous studies of the associations between BPV and risk of stroke, many of which exclusively included participants with hypertension, diabetes or CVD (Chapter 7). Secondly, primary care data have the potential to include much larger numbers of participants than possible previously (given the fact that measurements have already been taken in routine clinical practice). The absolute numbers of UK Biobank Welsh participants in whom visit-to-visit BPV could be measured in this study ranged from 2,352 to 4,658 (with ≥3 to ≥7 successive GP visits with coded SBP values, 5 years before recruitment). If these proportions were extrapolated to the UKB cohort as a whole (assuming the same data coverage and the same pattern of BP coding in the English and Scottish populations as in the UKB Welsh population), BPV could be measured in very large numbers (estimated range ~120,000 to ~200,000 with ≥3 to ≥7 coded SBP values, 5 years before recruitment in the whole UKB cohort). Advantages of this include firstly increased power to detect a potential association between BPV and risk of stroke. Secondly, it should mean that sufficient numbers of participants with sufficient outcomes will be available to detect potential differences in the associations between BPV and the main types and sub-types of stroke. Thirdly, it should be possible to conduct studies among participants with even more numbers of successive GP visits with coded SBP values (eg., ≥11 visits) Although these individuals were a relatively small proportion of the UKB Welsh population (eg. 14% with ≥11 GP visits with coded SBP values, 5 years

before recruitment), they would be relatively large group, in absolute numbers, when extrapolated to the whole UKB cohort (~70,000). In future, having these larger numbers of participants with different numbers of successive GP visits with coded BP values, over varying periods of time would enable sensitivity analyses of the influence of the number of SBP readings (and also the measurement periods) on the strength of the association between BPV and risk of stroke. This would potentially help to solve the unanswered question, 'how many BP values do we need, over what time periods, and at what frequency to most accurately predict an individual's BPV related stroke risk?'

A potential limitation of using coded healthcare data to measure BPV is the risk of selection bias. Individuals who have sufficient BP measurements to capture BPV most reliably (ie. those with the most BP measurements in any given time period) are likely to be those with the most variable BP. This might include individuals with hypertension (where variable BP may, in part, be due to effects of treatment and/or measurement error), and/or individuals with comorbidities which predispose to more variable BP (eg. individuals with pre-existing cardiovascular disease or renal disease). These factors should be accounted for in future analyses of the association between BPV and risk of stroke. In this study, I showed that selecting individuals with increasing numbers of GP visits with coded SBP values increased the prevalence of hypertension, diabetes, cardiovascular disease and renal disease. I also demonstrated that increasing the number of GP visits with coded SBP values increased the mean within individual BPV (mean within individual BPV increased from 10.7 to 13.6 when the number of visits increased from 3 to 11). Despite this, there was a sufficient range in BPV amongst included individuals, meaning that it would still be possible to demonstrate an association between increased BPV and risk of stroke. Provided confounding factors are accounted for, results from these selected populations should be internally valid, albeit less generalizable to the population as a whole.

A second potential limitation of using routinely collected coded healthcare data is that the number of BP readings taken, and time period over which they are recorded (including the potential gap between successive readings) is not the same for each

individual. Ideally, individuals should have the same number of BP readings selected over as similar a time period as possible to estimate BPV. Variation in these factors is likely to influence the precision of BPV measurement, and may also affect the strength of the observed association between BPV and risk of stroke (Chapter 7).

It is possible that some of the observed association between BPV and stroke is the result of residual confounding (from factors like renal failure). In future, access to primary care coded data could be used to construct additional variables for the presence/absence of renal disease amongst UK Biobank participants (eg. based on combinations of diagnostic Read codes and/or laboratory results). These could be used to adjust for these comorbidities more accurately than possible previously, enabling a less biased assessment of the association between BPV and stroke risk in UK Biobank.

8.5 Conclusions

I have shown that it is possible to estimate visit-to-visit blood pressure variability in large numbers of individuals using primary care coded data. In future, this type of data could be used to screen and select individuals, either for more detailed sub studies in UK Biobank (where different devices could be used to confirm BPV, like self-monitoring, or Ambulatory Blood Pressure Monitoring, ABPM), for other large prospective observational studies with linkages to primary care data, or for selection of participants in clinical trials. Ultimately, this type of data might be used in clinical practice to identify individuals with increased BPV, and therefore individuals at potentially increased risk of stroke due to BPV. I have shown that the BP values recorded in primary care can identify variation in BPV between individuals, and can therefore stratify individuals with higher vs. lower BPV. This is even possible (using ~3 BP readings) amongst individuals without a diagnosis of 'hypertension' and amongst individuals without conventional vascular risk factors. In the long run, if the associations between BPV and risk of stroke are confirmed, visit-to-visit BP measurements in primary care could be used to select individuals for more detailed monitoring, in the same way that individuals with high mean BP measurements are currently selected for home 24hr ABPM.

8.6 Appendix

Table 8.14 Selected Read codes for 'probable' and 'possible' hypertension

	'Probable hypertension'	**'Possible hypertension'**
Diagnosis codes	G2…including G20.. to G28../ G2y../G2z../G672./F4213./ F4042./L12../Gyu2./Gyu20/ Gyu21/G210./G2100./G2101./ G211./G2110/G2111/G21z./ G21z0/G21z1/G21zz/G220./ G221./G222./G22z./G230./ G231./G232./G233./G234./ G23z./L122./L1220/L1221/ LI223/L122z./L127./L127z/ L128./L1280/L1282/G240./ G2400/G240z/G241./G2410/ G241z/G244./G24z./G24z0/ G24z1/G24zz/	
Administration codes*		8HT5./67H8./8I3N./8BL0/ 8B26./8CR4./14A2./1JD../ 9N4L./9N03./9N1y2/ TJC7./TJC7z./U60C5/
BP monitoring codes*,†		9OI../9OI1./9OI2./9OI3./ 9O14./9OI5./9OI6./9OI7./ 9OI8./ 9OI9./9OIA./9OIZ./ 662../ 6624./6627./6629./ 662b./ 662c./ 6628./662F./ 662G./ 662H./662I./ 662O./662P./ 662P0/662d./ 6623./ 662V./662r./662j./

*Administration and BP monitoring codes which suggest a diagnosis of hypertension.

†I excluded the following BP monitoring codes ('BP reading raised' codes), because I didn't think they were sufficient to diagnose 'probable or possible hypertension' (ie.. one off high BP reading is not sufficient) 'R1y2.', '662B.', '662C.'

Table 8.15 Read codes which specify 'systolic' or 'diastolic' BP in their text definition

Specified systolic BP reading	
2469.	O/E systolic BP reading
246K.	Target systolic blood pressure
246N.	Standing systolic blood pressure
246Q.	Sitting systolic blood pressure
246S.	Lying systolic blood pressure
246j.	Systolic blood pressure centile
246e.	Ambulatory systolic BP
246W.	Average 24hr systolic BP
246d.	Average home systolic BP
246Y.	Average day interval systolic blood pressure
246b.	Average night interval systolic blood pressure

Specified diastolic BP reading	
246T.	Lying diastolic blood pressure
246X.	Average day interval diastolic blood pressure
246a.	Average night interval diastolic blood pressure
246L.	Target diastolic blood pressure
246A.	O/E diastolic BP reading
246R.	Sitting diastolic blood pressure
246f.	Ambulatory diastolic BP
246c.	Average home diastolic BP
246P.	Standing diastolic blood pressure
246i.	Diastolic blood pressure centile
246V.	Average 24hr diastolic BP

Table 8.16 Selection of data for Gender, Date of birth Townsend Index, and vascular comorbidities from the UK Biobank baseline assessment dataset.

Data field	Coding of data	Derived variables
Date of birth (33)	D/M/Y	
Gender (31)	0=female 1=male	
Townsend Index at recruitment (189)	Continuous numerical data	
Smoking status (20016)	0= never 1= previous 2= current	
Self-reported non-cancer diagnosis* (20016)	1065=hypertension, 1066=heart/cardiac problem 1067=peripheral vasc. disease 1072=essential hypertension 1073=gestational hypertension 1074=angina 1075=heart attack/MI 1076=heart failure 1077=arrhythmia 1078=heart murmur 1079=cardiomyopathy 1081=stroke 1082=TIA 1086=SAH 1087=intermittent claudication 1088=arterial embolism 1192=kidney/renal failure 1193=renal failure (dialysis) 1194=renal failure (no dialysis) 1220=diabetes 1222=type I diabetes 1223=type II diabetes 1471=atrial fibrillation 1491=brain haemorrhage 1583=ischaemic stroke 1607=diabetic nephropathy	Hypertension (if 'x' in field 20016= 1065, or 1072) Stroke (if 'x' in field 20016= 1081, 1086, 1491, or 1583) MI (if 'x' in the field 20016= 1075) Renal disease (if 'x' in the field 20016= 1192, 1193, or 1194) Diabetes (if 'x' in the field 20016= 1220, 1222, 1223, or 1607)

* From self-reported medical history, confirmed during a brief nurse-led interview at the UK Biobank baseline assessment.

Table 8.17(A to E) Frequencies of individual BP-related Read codes.*

A) Hypertension diagnosis codes

Code	Read code Text definition	Code frequency
G2…	Hypertensive disease	2,074
G20..	Essential hypertension	3,148
G200.	Malignant essential hypertension	2
G201.	Benign essential hypertension	76
G202.	Systolic hypertension	10
G203.	Diastolic hypertension	0
G20z.	Essential hypertension NOS	155
G21..	Hypertensive heart disease	4
G210.	Malignant hypertensive heart disease	0
G2100	Malignant hypertensive heart disease without CCF	0
G2101	Malignant hypertensive heart disease with CCF	0
G211.	Benign hypertensive heart disease	0
G2110	Benign hypertensive heart disease without CCF	0
G2111	Benign hypertensive heart disease with CCF	0
G21z.	Hypertensive heart disease NOS	0
G21z0	Hypertensive heart disease without CCF	0
G21z1	Hypertensive heart disease with CCF	1
G21zz	Hypertensive heart disease NOS	2
G22..	Hypertensive renal disease	2
G220.	Malignant hypertensive renal disease	1
G221.	Benign hypertensive renal disease	0
G222.	Hypertensive renal disease with renal failure	0
G22z.	Hypertensive renal disease not otherwise specified	3
G23..	Hypertensive heart and renal disease	0
G230.	Malignant hypertensive heart and renal disease	0
G231.	Benign hypertensive heart and renal disease	0
G232.	Hypertensive heart and renal disease with CCF	0
G233.	Hypertensive heart and renal disease with renal failure	0
G234.	Hypertensive heart and renal disease with renal failure and CCF	0
G23z.	Hypertensive heart and renal disease NOS	0
G24..	Secondary hypertension	3
G240.	Secondary malignant hypertension	0
G2400	Secondary malignant renovascular hypertension	0
G240z	Secondary malignant hypertension NOS	0
G241.	Secondary benign hypertension	0
G2410	Secondary benign renovascular hypertension	0
G241z	Secondary benign hypertension NOS	0
G244.	Hypertension secondary to endocrine disorders	0
G24z.	Secondary hypertension NOS	0
G24z0	Secondary renovascular hypertension NOS	0
G24z1	Hypertension secondary to drugs	0
G24zz	Secondary hypertension NOS	1
G25..	Stage 1 hypertension (NICE 2011)	0
G26..	Severe hypertension (NICE 2011)	0
G27..	Hypertension resistant to drug therapy	0
G28..	Stage 2 hypertension (NICE 2011)	0
G2y..	Other specified hypertensive disease	1
G2z..	Hypertensive disease NOS	24
G672.	Hypertensive encephalopathy	0
F4213	Hypertensive retinopathy	12
F4042	Blind hypertensive eye	0
L12..	Hypertension complicating pregnancy	5
L122.	Other pre-existing hypertension in pregnancy/childbirth	0
L1220	Other pre-existing hypertension in pregnancy/childbirth	0
L1221	unspecified	0
L1223	Other pre-existing hypertension in pregnancy/childbirth delivered	0

L122z	Other pre-existing hypertension in pregnancy/childbirth not	0
L127.	delivered	0
L127z.	Other pre-existing hypertension in pregnancy/childbirth NOS	0
L128.	Pre-eclampsia or eclampsia with pre-existing hypertension	0
L1280	Pre-eclampsia or eclampsia with pre-existing hypertension NOS	0
L1282	Pre-existing hypertension complicating pregnancy	0
	Pre-existing hypertensive heart disease complicating pregnancy	
	Pre-existing secondary hypertension complicating pregnancy	
Gyu2.	Hypertensive diseases	0
Gyu20	Other secondary hypertension	0
Gyu21	Hypertension secondary to other renal disorders	0
61462	Hypertension induced by oral contraceptive pill	0
Total		**5,524**

B) Hypertension administration codes

Code	Definition	Code frequency
8HT5.	Referral to hypertension clinic	1
67H8.	Lifestyle advice regarding hypertension	285
8I3N.	Hypertension treatment refused	1
8BL0.	Patient on maximal tolerated HTN treatment	72
8B26.	Antihypertensive therapy	2
8CR4.	Hypertension clinical management plan	0
14A2.	History of hypertension	180
1JD..	Suspected hypertension	22
9N4L.	DNA. Hypertension clinic	5
9N03.	Seen in hypertension clinic	256
9N1y2	Seen in hypertension clinic	291
TJC7.	Adverse reaction to other antihypertensives	0
TJC7z	Adverse reaction to hypertensives NOS	0
U60C5	Adverse reaction to other antihypertensive drugs	0
Total		**1,115**

C) Hypotension diagnosis codes

Code	Definition	Code frequency
G87.	Hypotension	18
G870.	Orthostatic hypotension	60
G871.	Chronic hypotension	0
G872.	Idiopathic hypotension	0
G873.	Hypotension due to drugs	0
G87z.	Hypotension NOS	1
Total		**79**

D) Blood pressure 'monitoring' codes

Code	Definition	Code frequency
9OD..	Hypertension screen admin	171
9OD1.	BP screen -1st call	124
9OD2.	BP screen- 2nd call	87
9OD3.	BP screen – 3rd call	46
9OD4.	BP screen- call deleted	0
9OD5.	BP screen – 1st recall	6
9OD6.	BP screen- 2nd recall	2
9OD7.	BP screen- 3rd recall	7
9OD8.	BP screen- recall deleted	0
9OD9.	BP abnormal – 1st recall	7
9ODA.	BP abnormal – 2nd recall	0
9ODB.	BP abnormal- 3rd recall	0
9OI..	Hypertension monitoring admin	171
9OI1.	Attends hypertension monitoring	0
9OI2.	Refuses hypertension monitoring	0
9OI3.	Hypertension monitoring offer default	0
9OI4.	Hypertension monitoring- 1st letter	728
9OI5.	Hypertension monitoring- 2nd letter	129
9OI6.	Hypertension monitoring – 3rd letter	45
9OI7.	Hypertension monitoring – verbal interview	54
9OI8.	Hypertension monitoring – phone invitation	39
9OI9.	Hypertension monitoring deleted	0
9OIA.	Hypertension monitoring check done	623
9OIZ.	Hypertension monitoring admin NOS.	0
68B1.	Hypertension screen	646
68B4.	Risk factors present at hypertension screen	12
ZV70B	Examination of blood pressure	79
ZV7B1	Screening for hypertension	39
8HRH.	Referral for ambulatory BP monitoring	24
8IBA.	Ambulatory BP monitoring not indicated	0
8I3Y.	Blood pressure procedure refused	9
315B.	Ambulatory BP recording	275
662..	Hypertension monitoring	5,739
6624.	Borderline hypertension- yearly observations	11
6627.	Good hypertension control	152
6629.	Hypertension follow-up default	1
662b.	Moderate hypertension control	4
662c.	Hypertension 6 month review	224
662L.	24hr BP monitoring	157
6628.	Poor hypertension control	35
662F.	Hypertension treatment started	43
662G.	Hypertension treatment changed	131
662H.	Hypertension treatment stopped	9
662I.	No record of BP reading	11
662O.	Good hypertension control	22
662P.	Hypertension monitoring	2,659
662P0	Hypertension 9 month review	1
662d.	Hypertension annual review	1,078
662Q.	Borderline BP	54
6623.	Pre-treatment BP reading	53
662V.	Blood pressure monitoring	983
662j	Blood pressure recorded by patient at home	56
662r.	Trial withdrawal of antihypertensive therapy	0
662B.	O/E- initial high BP	5
662C.	O/E- check high BP	5

R1y3	Low BP reading	9
R1y2	Raised BP reading	705
R1y4	BP reading labile	13
Total		**15,483**

E) Blood pressure value codes

Code	Definition	Code frequency
246..	O/E blood pressure reading	144,868
2469.	O/E systolic BP reading	369
246A.	O/E diastolic BP reading	343
246C.	Lying BP reading	3
246D.	Standing BP reading	5
246E.	Sitting BP reading	69
246F.	O/E BP decreased	1
246J.	O/E BP reading- no postural drop	1
2468.	O/E BP reading- postural drop	47
246K.	Target systolic blood pressure	85
246L.	Target diastolic blood pressure	98
246N.	Standing systolic blood pressure	0
246P.	Standing diastolic blood pressure	1
246Q.	Sitting systolic blood pressure	4
246R.	Sitting diastolic blood pressure	0
246S.	Lying systolic blood pressure	0
246T.	Lying diastolic blood pressure	0
246Z.	O/E blood pressure reading NOS	211
246i.	Diastolic blood pressure centile	0
246j.	Systolic blood pressure centile	0
246X.	Average day interval diastolic blood pressure	3
246Y.	Average day interval systolic blood pressure	6
246a.	Average night interval diastolic blood pressure	2
246b.	Average night interval systolic blood pressure	5
2460.	BP unrecordable	0
2465.	BP reading borderline raised	361
2466.	BP reading raised	496
2467.	BP reading very high	9
2461.	BP reading very low	1
2462.	BP reading low	7
2463.	BP borderline low	2
2464.	O/E- BP normal	1,790
246B.	O/E – BP stable	2
246G.	O/E – BP labile	2
246M.	White coat hypertension	43
246V.	Average 24hr diastolic BP	4
246W.	Average 24hr systolic BP	6
246e.	Ambulatory systolic BP	2
246f.	Ambulatory diastolic BP	1
246c.	Average home diastolic BP	0
246d.	Average home systolic BP	6
246g.	Self-measured BP readings	1
Total		**148,941**

*Amongst 171,142 BP- related Read codes in the Welsh coded primary care dataset. Shaded rows indicate Read codes which were never used.

Table 8.18 Distribution of within-participant median gaps of ≤ 18 months

Gap time (days)	Participants (n)	% (of 5,050)*	% (of 7,748)[†]
≥1 to ≤30	415	8.2	5.4
>30 to ≤60	619	12.3	7.9
>60 to ≤90	501	9.9	6.5
>90 to ≤120	404	8.0	5.2
>120 to ≤150	324	6.4	4.2
>150 to ≤180	440	8.7	5.7
>180 to ≤210	335	6.6	4.3
>210 to ≤240	248	4.9	3.2
>240 to ≤270	217	4.3	2.8
>270 to ≤300	209	4.1	2.7
>300 to ≤330	203	4.0	2.6
>330 to ≤365	255	5.0	3.3
>365 to ≤395	193	3.8	2.5
>396 to ≤425	153	3.0	1.9
>426 to ≤455	151	2.9	1.9
>456 to ≤485	135	2.7	1.7
>486 to ≤515	122	2.4	1.6
>516 to ≤550	126	2.5	1.6
Total	5050	100	100

*Total number of individuals who had a median within-participant gap of ≤ 18 months

[†]Total number of individuals who have at least 2 GP visits with coded SBP values prior to UK Biobank recruitment.

Chapter 9 Conclusions and future directions

9.1 Main findings of thesis

I have used linkages to multiple different sources of routinely collected, coded healthcare data to detect and quantify a potential exposure, blood pressure variability (BPV), and to optimise the ascertainment and confirmation of an important clinical outcome, stroke, in a sub-cohort of the UK Biobank population. My work will inform the design of future nested case cohort or case control studies of the associations between BPV and risk of stroke in UK Biobank.

9.1.1 Accuracy of coded healthcare data for identifying stroke cases in UK Biobank

My objective in Chapter 2 was to perform a systematic review of the accuracy of routinely available coded healthcare data for stroke and its main pathological types, to inform the selection of codes for the identification of stroke cases in UK Biobank.

Among 37 studies, the accuracy (PPV) of coded hospital or death certificate data ranged from poor to excellent, but studies varied widely in their settings, methods, reporting, quality, and in the choice and accuracy of codes. Appropriately selected, 'stroke specific' codes appeared sufficiently accurate to identify stroke cases in UK Biobank. Broad cerebrovascular codes were consistently less accurate for stroke, but these codes may be used to identify additional potential cases when further confirmation steps are planned. Few studies assessed either coded primary care data or combinations of data sources for stroke.

9.1.2 Accuracy of patient self-report of stroke

My objective in Chapter 3 was to perform a systematic review of the accuracy of patient self-report of stroke, to inform the use of participant self-report data for the identification of stroke cases in UK Biobank.

Among 17 studies, the accuracy (PPV) of patient self-report of stroke, ranged from poor to very good. As expected, accuracy of self-report increased in populations with higher stroke prevalence. Amongst the studies with low stroke prevalence, like UK Biobank, between a 1/3 and 3/4 of self-reported strokes were false positive. I

concluded that self-report is unlikely to be helpful for identifying stroke cases in UK Biobank, without subsequent confirmation. Self-report may be useful for case ascertainment in combination with other data sources.

9.1.3 Reliability and feasibility of ischaemic stroke classification systems

My objective in Chapter 4 was to perform a systematic review of the inter- and intra-observer reliability of ischaemic stroke classification systems, and amongst included studies, to report the proportion of cases classified to a single determined subtype, to inform the approaches to ischaemic stroke classification in UK Biobank.

Amongst 25 studies which assessed the overall inter-observer reliability of one or more of the existing ischaemic stroke classification systems, reliability ranged from moderate to excellent. Characteristics other than the classification system used contributed much of the variation in reliability. Use of clear rules, data abstraction protocols, computer based classification and fewer categories all improved reliability. There was insufficient evidence to recommend the newer 'single cause mechanistic' classification systems in place of older, more established systems. The proportion of cases undetermined to a single subtype was lowest using 'anatomical' or 'descriptive mechanistic' classification systems. I concluded that no single classification system was fit for every purpose. I recommended a flexible approach to ischaemic stroke classification in UK Biobank, along with the features (above) which have been shown to enhance reliability.

9.1.4 Choosing Read codes to identify stroke cases in UK Biobank

My objective in Chapter 5 was to select Read codes to identify 'acute' stroke cases in UK Biobank, for future nested case cohort or case control studies which require high accuracy (PPV) for stroke, and Read codes to identify additional 'prevalent stroke cases', for exclusion of participants from future studies of 'first in a lifetime' stroke.

I searched multiple online resources, and selected 57 Read codes for ‘acute stroke’, from among >200 potential codes. No single online resource included all of the relevant codes for stroke. I hypothesised that my ‘acute stroke’ codes would have high accuracy (high PPV) for stroke, based on matching to previously validated ‘stroke specific’ International Classification of Diseases (ICD) codes (Chapter 2). I identified an additional 18 Read codes for possible ‘past history of stroke’, which could be used to identify additional ‘prevalent stroke’ cases, and exclude participants from future studies of ‘first in a lifetime’ stroke. Read codes matching the broader cerebrovascular diseases ICD codes could also be used for this purpose. I concluded that further studies are required to validate groups of Read codes for stroke and its main pathological types, and to compare the contribution of multiple overlapping data sources for stroke.

9.1.5 Identification of prevalent and early incident stroke cases in UK Biobank

My objective in Chapter 6 was to identify prevalent and early incident stroke cases in UK Biobank, and in a sub-cohort of ~22,000 participants, using multiple overlapping coded data sources, to determine to proportion of cases identified by each individual source, the added contribution of linkage to coded primary care data, and the number of cases classified to main pathological stroke type.

The majority of prevalent stroke cases in UK Biobank (~9,000 in the whole cohort and ~1,000 in the sub-cohort) were identified by participant self-report. Most of these cases were ‘unspecified’ pathological type. By contrast, the 850 incident cases identified by coded hospital and/or death certificate data in the whole UK Biobank population were mostly ‘specified’ pathological type. In the sub-cohort with complete linkage to all three coded data sources, just over a third of ~180 incident cases were detected using coded primary care data, and half of the of these were not detected by any other data source. I concluded that, as expected, linkage to coded primary care data should improve the completeness of stroke outcomes identification in UK Biobank, and that the majority of cases detected using coded hospital or death certificate data may be classified into main pathological type, without further confirmation.

9.1.6 The association between blood pressure variability and risk of stroke

My objective in Chapter 7 was to review published studies of the association between medium- to long-term blood pressure variability and risk of stroke, to determine the strength of the existing evidence, limitations of previous studies, and to inform potential approaches to future BPV measurement amongst UK Biobank participants.

All sixteen studies demonstrated a trend towards an association between BPV and stroke risk, but in many, small numbers of stroke outcomes limited the ability to detect a significant association, and the effect size was often small. The majority of studies were in selected populations with increased vascular risk. The strength of the association between BPV and stroke appeared to increase with the precision of BPV measurement, but there was no consensus on how BPV should be measured, including number of BP readings and/or time period over which readings were recorded, to best predict stroke risk. I concluded that there was definite evidence that BPV is an important potential risk factor for stroke, but future, large, population based studies are required to reliably test this association, and to explore the optimum method(s) of BPV measurement for potential future BP monitoring and treatment.

9.1.7 Can coded primary care data be used to measure blood pressure variability?

My objective in Chapter 8 was to investigate the potential of using routinely collected coded primary care data to estimate visit-to-visit BP variability (BPV) in a sub-cohort of ~10,000 UK Biobank participants.

It was possible to estimate visit-to-visit BPV, using ≥3 GP visits with coded BP values before recruitment, in the majority of UK Biobank participants. Using an exposure period of 5 years it was possible to estimate visit-to-visit BPV in just under half, but these participants had a higher prevalence of hypertension, diabetes, and cardiovascular disease compared to the baseline UKB Welsh population. Although selecting participants reduced generalisability, there was good variability in BP variability amongst those selected (SD SBP ranged from ~5 to ~7mmHg), and there

was reasonable agreement between mean coded BP and an independent reference standard. I concluded that routinely collected coded primary care data should enable exploration of the associations between BPV and stroke in a 'real world setting' and in much larger numbers than previously possible.

9.2 Systematic review methodology

The systematic reviews in this study differed from conventional methodology in a number of ways.

Firstly, study protocols were not registered or published in advance of the reviews taking place. This is now a requirement for systematic reviews to ensure that research methods are transparent and to avoid 'data driven research', where study methods (including inclusion/exclusion criteria and/or subgroup analyses) are influenced by the availability of published data and/or 'positive' outcomes.(The PloS Medicine Editors 2011) Registration of the study protocol should also reduce publication bias, which is an increasingly recognised problem in systematic reviews. My systematic review of ischaemic stroke classification systems was 'data-driven', because I focussed on the kappa statistic only after discovering that it was the most widely used and reported measure of the inter-rater reliability of ischaemic stroke classification systems. I will have excluded studies which used different statistical measures of reliability, and these studies although in the minority, may have led me to different conclusions had they been included.

Secondly, it is standard protocol in systematic reviews that all titles, abstracts, and full texts for inclusion are independently reviewed by two researchers. This process minimises random error and bias in study selection. In my systematic reviews, only a proportion of potential studies were reviewed for inclusion by a second, independent researcher. This means that it is more possible that I excluded relevant studies by mistake, or included irrelevant studies (and potentially positive studies) in a biased fashion. Use of a second reviewer also ensures that study inclusion/exclusion criteria are easy to follow and lead to reproducible results.

Thirdly, in some of my systematic reviews I excluded studies with less than 50 included cases. Although I may have excluded important and relevant studies by doing this, smaller studies may be more prone to publication bias, and by excluding these studies I am hoping to have minimised the effect of publication bias on my results.

9.3 Advantages of using routinely collected coded healthcare data for research in large populations

I have illustrated multiple advantages of using routinely collected coded healthcare data for research in large populations.

Describing exposures in large populations and enhancing statistical power to detect exposure-outcome relationships.

Firstly, linkage to coded data facilitated the measurement of a novel exposure, BPV, in a very large population. I found that BPV could be measured, using at least 3 separate GP visits within 5 years of recruitment, in around 4,600 UK Biobank participants using coded primary care data (~47% of a sub-cohort of ~10,000). Scaled to the whole UK Biobank population of around 500,000, the numbers of participants in whom BPV could be estimated through primary care data linkages, (potentially up to ~230,000), surpasses those included in the majority of individual previous clinical trials or observational studies of associations between BPV and risk of stroke. So far, the largest individual population based study of the association between BPV and stroke included ~120,000 participants, all of whom had a diagnosis of hypertension.(Yu et al. 2014)

As well as very large sample size, the added advantage of UKB is the unparalleled depth and breadth of phenotype and genotype data collected at baseline, and the strategy (in development) for more complete and accurate stroke case identification, confirmation and sub-classification during follow-up. By contrast, the previous study, mentioned above, ascertained stroke cases using less accurate ICD codes (including codes for ‘cerebrovascular diseases’), without subsequent confirmation. (Yu et al. 2014). It is therefore likely to have included larger numbers of false positive cases during follow-up, thereby potentially underestimating the strength of the observed association between BPV and risk of stroke.

In the long run, combining UK Biobank data with data collected in other large cohort studies, e.g., the China Kadoorie Biobank (CKB),(Chen et al. 2005) the Prospective Studies Collaboration (PSC),(Prospective Studies Collaboration et al. 2002) the Million Women Study (MWS),(The Million Women Study Collaboration. 1999) and others, will increase included numbers further, and further increase statistical power. Comparing data from these different populations, for example China versus UK, and/or other countries worldwide, will enable regional differences in the associations between BPV, the risk of stroke, its main types and subtypes to be explored.

More representative than previous studies of the associations between BPV and risk of stroke

Secondly, coded primary care data appeared to be more representative of the general population than data gained from highly selected clinical trial populations, where measurements were taken in a controlled environment, and were therefore less representative of the 'real world' setting. I showed that the UKB participants who had multiple successive BP values in the coded primary care dataset had fewer vascular risk factors compared to participants included in the majority of previous studies of the association between BPV and risk of stroke. Furthermore, the selection of participants using routinely collected coded data did not limit the potential to detect a true association between BPV and increased risk of stroke. Despite the fact that individuals with large numbers of BP measurements had higher BPV overall than those with fewer BP measurements, I found that there was a wide range in variability in BPV amongst these individuals, even among those with >11 successive BP readings.

Depth and breadth of data for well-powered and detailed sub-group analyses, enabling personalised and stratified risk prediction.

Thirdly, data collected and coded in routine clinical practice were wider ranging than data collected de-novo for clinical trials and observational studies (where time and cost often limits the amount of data collected). I found a wide range in the number of successive BP measurements amongst UKB participants (from 2 to >20), over a wide range of durations (periods of >1 to >10 years) using linkages to coded primary care data. In future, these coded data could be used to conduct sensitivity analyses to compare the influence of the number of BP readings used, and time period over

which readings were taken (precision of BPV measurement), on stroke risk prediction. We may, therefore, be more likely to be able to answer the question, 'How many BP measurements, over what time periods, and at what frequency, do we need to most accurately predict BPV related stroke risk?', by analyses of coded data collected as part of routine practice, in addition to or instead of data collected specifically for prospective observational studies or clinical trials.

A recent published study used coded electronic health record data from primary care, hospital admissions, death certificates, and the Myocardial Ischaemia National Audit Project (MINAP), to investigate associations between mean blood pressure and cardiovascular outcomes, including stroke.(Rapsomaniki et al. 2014) The study demonstrated the increased risk of stroke associated with a 20mmHg increase in mean blood pressure, stratified by age group, and by different blood pressure values in each age group. Around 65% of the cohort (~1,300,000 of ~2,000,000) had at least one coded blood pressure value recorded in the primary care dataset within 2 years of recruitment to the study (compared with 57% of participants within 2 years of recruitment to UK Biobank). In agreement with my work, participants in this study with no coded BP values were more likely to be male and were healthier. In comparison to previous large studies of the associations between mean BP, stroke and its main pathological types, (PSC and APCSC) the above study demonstrated the association between mean BP and stroke across a wider range of blood pressures, and identified, for the first time, a nadir below which reducing mean BP did not reduce risk of stroke further (~130mmHg systolic and ~80mmHg diastolic BP). It provides an example of how including a wider range of exposures, which is possible at scale using linkages to coded healthcare data, can provide new and valuable information about a previously well studied exposure-outcome relationship.

Although selection of individuals using coded data limited generalisability to the population as a whole (these individuals had increased BPV and an increased prevalence of vascular risk factors compared to the general population), future sub-group analyses within these populations could be translated to sub-groups of the general population for personalised and stratified risk prediction. Personalised and stratified medicine is a major goal of current clinical research. Powerful sub-group

analyses exploring the influence of age, and a variety of other baseline characteristics on the associations between BPV and stroke should be possible in UK Biobank, as I have already demonstrated that large numbers of participants will have sufficient measures of exposures (BPV estimates), and, in time, large numbers of outcomes (incident stroke and/or its main pathological types) should accrue.

Linked coded data should also facilitate adjustment for more confounding factors than possible in previous studies, for example, renal impairment, which is an important potential confounder of the association between BPV and risk of stroke. The opportunity to link to the UKB baseline assessment, which includes a wide range of prospectively collected data on lifestyle, environment, co-morbidities, and medication use, further broadens the scope of the analyses possible in UK Biobank.

Accuracy of coded data

Finally, I have shown that using linked coded data for defining exposures, or for follow-up of disease outcomes, does not necessarily reduce accuracy. I showed good agreement between mean BP estimated using the coded primary care dataset and mean BP measured at the UK Biobank baseline assessment. The majority of stroke cases identified using ICD coded data could be classified into a main pathological type of stroke (up to 89% of stroke cases identified using ICD codes based on hospital admissions and death certificates during follow-up until December 2010), and these codes had PPVs >90% for either ischaemic or haemorrhagic stroke (based on a systematic review of the published data).

The use of multiple overlapping data sources is likely to improve the completeness of stroke case detection, and reduce selection bias. I showed that using multiple sources of linked coded data (adding coded primary care data to ICD codes from hospital admissions and death certificates) increased the number of stroke cases detected during follow-up in a subset of UK Biobank participants to the end of December 2010. Around 19% of these cases would not have been detected without the addition of coded primary care data, and it is likely that missing these cases would under represent milder subtypes of ischaemic stroke (e.g. lacunar ischaemic stroke). A similar advantage of using multiple sources of routinely collected healthcare data was demonstrated in a study published in 2010. This showed that the incidence of

myocardial infarction was underestimated by 25 to 30% if only one source of coded data was used for case ascertainment, compared to multiple overlapping sources (primary care, hospital admissions, death certificates, and the MINAP disease registry).(Herrett et al. 2013)

9.4 Future directions

Challenges of using routinely collected coded-healthcare data for research

Routine, coded healthcare data are not collected primarily for research purposes. Inconsistencies between data from different geographical areas and/or different healthcare settings need to be interrogated, understood and resolved before these data can be used in research studies such as UK Biobank. For example, data from primary care and secondary care are coded differently (Read versus ICD), and different versions of Read are currently in use in England, Scotland and Wales, depending on the particular GP practice systems used. Furthermore, data from different regions may have variable completeness, and the periods of data coverage may vary between region and/or data source. I rationalised these inconsistencies in my research by censoring data so that my denominator population (9,947 in the Welsh UKB sub-cohort) had as complete coverage as possible from all three data sources: primary care, hospital admissions, and death certificates. However, linkages to coded primary care data were only ~50% complete in Wales for UK Biobank at the time of this study. In future, it is hoped that linkages will approach ~100% completeness for all UKB participants and across all three data sources, across England, Scotland, and Wales.

I have found that creating code lists to define exposures, like BPV and outcomes, like stroke, is a time consuming process. The codes selected to represent these variables will depend on the research question (for example a distinction was made between ‘incident’ versus ‘prevalent’ stroke, and different groups of codes might be chosen to maximise sensitivity and/or maximise PPV of stroke ascertainment/confirmation). Eventually, groups of codes combined in algorithms used to define health related outcomes such as stroke, as well as those for additional exposures such as blood pressure and blood pressure variability, will be provided via

the UK Biobank Data Showcase website for sharing with the wider research community.(http://www.ukbiobank.ac.uk/) It is hoped that sharing case definitions will increase transparency of research using coded healthcare data, increase efficiency (because code lists do not have to be recreated) and improve reproducibility. In future, analyses in UK Biobank might investigate the influence of different groups of codes selected (e.g. 'stroke specific codes' chosen to maximise PPV versus more general cerebrovascular codes chosen to maximise sensitivity) on the strength, shape and precision of the associations between mean BP (and or BPV) and risk of stroke, its main types, and subtypes.

Validating BPV estimates and potential implications for future research and clinical practice

Although there was good within participant agreement between mean BP in primary care data and BP recorded at the UKB baseline assessment, I have not been able to validate my estimates of pressure variability (BPV). In future, I aim to determine the association between within participant BPV estimated using coded primary care data, and known correlates of BPV (e.g. renal impairment, and/or age), using linkages to the UK Biobank baseline assessment dataset and/or coded primary care data. In collaboration with relevant experts, my supervisor's group is now conducting research to create similar algorithms for renal impairment in the UK Biobank Welsh sub-cohort, using coded primary care data for renal function (i.e. coded values of measures such as creatinine and glomerular filtration rate), and/or coded diagnoses of renal disease. Alternatively, coded primary care data might be used to select a subset of participants with a range of BPV values for comparison with direct, accurate estimates of BPV, e.g. successive home-monitoring, or successive Ambulatory Blood Pressure Monitoring (ABPM), enabling calibration of the primary care based BPV estimates.

In future, coded primary care data may not be sufficiently accurate to predict an individual's BPV related stroke risk, but they may be used to screen for potential BPV (i.e. to detect UKB participants with potential increased BPV). Coded data might therefore be used for BPV ascertainment, and more accurate methods could be used for subsequent confirmation and risk stratification. This approach might be applied in UK Biobank sub-studies, in clinical trials, or in future clinical practice

(analogous to the use of outpatient BP screening in primary care to select individuals for 24hr ABPM prior to confirmation and treatment of hypertension).

A screening test requires high sensitivity, and a limitation of using coded primary care data to detect individuals with/without potential BPV in a real world setting, is selection bias. Individuals without BP measurements and/or with insufficient measurements for accurate BP estimation would be missed by any screening process based on routinely collected coded data. Although the majority of UK Biobank participants in the Welsh sub-cohort had at least one BP value, I showed that around 40% did not have ≥ 3 BP values within 10 years of recruitment, and that individuals with multiple successive BP readings had a higher prevalence of vascular risk factors. Importantly, individuals with fewer vascular risk factors who were less likely to have multiple successive BP values in the primary care dataset were also likely to benefit most from BPV screening and treatment. Individuals with fewer vascular risk factors demonstrated a stronger association between BPV and risk of stroke in previous studies.(Rothwell et al. 2010b) Therefore, if the association between BPV and risk of stroke is confirmed in future, it would be important to highlight this potential 'at risk' population. Future public health programmes might focus on increasing BP screening in primary care, particularly amongst populations with low vascular risk. To facilitate this, future research in UK Biobank could investigate the minimum number of successive BP values, over the longest period of time, which reproducibly estimates within individual BPV. This would minimise the additional burden on GP workload, and increase the number of potential individuals in whom BPV could be estimated.

Developing and validating an algorithm for the identification, confirmation and sub-classification of stroke outcomes in UK Biobank

I showed that ICD coded data classified ~89% of stroke cases into ischaemic or haemorrhagic stroke in the UK Biobank Welsh sub-cohort. Although these codes had PPVs of ≥90% for the main pathological types of stroke in previous published studies, they will need to be validated against an independent reference standard to determine their accuracy for stroke and its main types in the UK Biobank population. Classification should minimise inclusion of false positive cases, which reduces power to detect differences in the strength of the associations between BPV and ischaemic or haemorrhagic stroke, or between BPV and their main subtypes.

A sub-study is now ongoing in my supervisor's group to validate a selection of potential stroke cases ascertained by ICD codes from hospital admissions/death certificates, Read codes, participant self-report, and/or combinations of these data sources amongst Scottish UK Biobank participants. A panel of local experts will determine the reference standard 'stroke' versus 'non stroke', and the pathological types and subtypes of stroke, where applicable, using access to all available medical record data, blinded to the coded diagnosis. This will enable estimation of the accuracy (PPV) of selected codes for stroke, and its main pathological types in UK Biobank. In addition, linkages to the Scottish Stroke Care Audit (SSCA), which collects data on all strokes diagnosed in Scottish hospitals, including those seen in outpatient clinics (representing a proportion of non fatal, non hospitalised strokes), could be used to estimate the sensitivity of linked, coded data for stroke. However, a true population based reference standard (including all hospitalised, non hospitalised and fatal strokes) would need to be used to estimate the true sensitivity and PPV of these data sources in UK Biobank.

I showed that there were insufficient published data on the accuracy of Read codes from primary care, or of multiple overlapping coded data sources for stroke, or its main types. The validation work (above) should provide the first results of this type, and may determine the potential of coded primary care data to improve the confirmation and sub-classification of otherwise unspecified stroke cases identified using hospital admissions data. For example, a proportion of the ~11% of cases which were of unspecified pathological type in the UK Biobank population (based

on ICD codes alone) may be classified into main pathological types of stroke by cross linking to Read codes for ischaemic or haemorrhagic stroke.

It seems unlikely that coded data alone will be sufficient to sub-classify detailed pathological subtypes of stroke (i.e. subtypes of ischaemic stroke and of subarachnoid and intracerebral haemorrhage) during follow-up in UK Biobank. Additional medical record data will be required for this, but the approach should remain efficient and scalable across the whole UK Biobank cohort. Within my supervisor's group, an algorithm is in development in a sub-cohort of the UKB population (NHS Lothian) to guide confirmation and classification of potential stroke cases identified by ICD coded data, using unstructured data extracted from electronic medical records. In the first phase of this work, UKB participants identified using ICD codes for stroke admissions were linked to the inpatient electronic healthcare record, which included inpatient hospital discharge letters, outpatient clinic letters, and results of tests (including reports of CT scans). This information was used to classify 98% (95% CI 95% to100%) of confirmed stroke cases into a main pathological type, 88% (95% CI 78% to 98%) of confirmed ischaemic strokes into ischaemic stroke subtypes using the OCSP anatomical ischaemic subtype classification, and 39% (95% CI 13% to 67%) of confirmed ischaemic strokes into ischaemic stroke subtypes using the TOAST mechanistic sub-classification. Subsequent phases of this work will extend this process to non hospital admitted stroke cases, and consider ways in which an algorithm for confirmation and sub-classification of stroke cases could be applied at scale throughout the UK where UK Biobank participants were recruited, live, and receive healthcare. To enable the scalability required, adjudication of stroke cases during follow-up in UK Biobank may be possible via a web-platform, using data extracted by a distributed network of clinical research staff (e.g. those engaged with the UK Clinical Research Network) from hospital and primary care unstructured electronic (or if necessary paper) medical records. Once these data are anonymised and uploaded, volunteers (stroke physicians and/or stroke trainees), could use this web-based platform to confirm potential stroke cases, and classify them into their main pathological types and subtypes. To maximise reliability, data should be abstracted from medical records following written protocols, training cases could also be included online, and

classification could be driven by a standardised computer-based process. I demonstrated that these approaches improved the inter-observer reliability of ischaemic stroke classification. Furthermore, data used for classification would be retained on the platform, enabling different combinations of outcomes, including intermediate disease phenotypes (like the presence/absence of large artery atherosclerosis, or small vessel disease) to be investigated in future. This is similar to the 'phenotypic classification' approach developed for the A-S-C-O and phenotypic CCS ischaemic stroke classification systems.

Investigating associations between BPV and the main types and subtypes of stroke in UK Biobank

It is not yet known whether blood pressure variability associates more strongly with haemorrhagic stroke than ischaemic stroke, or if differences exist in the strength of associations between BPV and the main pathological subtypes of stroke. Large numbers of classified cases are needed to reliably investigate these potential differences. In this research I have demonstrated that linkages to multiple sources of routinely collected coded healthcare data provide an exciting opportunity to begin to explore these research questions in a very large population. Having developed methods to define both exposures and outcomes in UK Biobank, I hope in the future to use this in depth methodological understanding to investigate in the UK Biobank population the associations of visit-to-visit blood pressure variability with incident stroke and its main pathological types, adjusting for changes in mean BP over time, and a range of potential confounding factors. In the long run, this work should improve our understanding of stroke type/subtype specific aetiology, and will potentially provide future targets for drug development in clinical trials (specifically treatment of BPV, which may provide new avenues for personalised and stratified stroke risk prevention).

Wider implications of the importance of BPV

Blood pressure variability may be a new risk factor for stroke, with implications for monitoring and treatment beyond mean BP. Individuals with normal mean BP and high BP variability, (or normal mean BP and episodic hypertension) would not otherwise be considered 'at risk' and would not otherwise receive BP monitoring or treatment. The previously observed association between BPV and stroke is highest

amongst those with otherwise normal mean BP.(Rothwell et al. 2010b) Therefore, these individuals have the most potential to gain from new treatments targeted at BPV reduction. In addition, it is possible that BPV is one of the contributing factors in 'cryptogenic stroke', stroke without any clear aetiology, which can affect up to a third of 'young stroke' patients. Recognition of BPV as a potential risk factor may lead to a step forward in the prevention of cryptogenic stroke.

Measurement of BPV is likely to be time consuming and costly in the 'real world'. This effort may add little to the treatment of individuals with BPV who are already hypertensive (for example, these individuals could receive an antihypertensive drug chosen to reduce both mean BP and BPV, see below). However, prevention of stroke in younger patients (as described above) is likely to have the most long-term economic impact (by reducing the longer term care burden). Future clinical trials of BPV reduction may have more impact if they focus on the identification and treatment of these groups of individuals.

Post hoc analyses of clinical trials have demonstrated that certain antihypertensive medications reduce BPV more than others. In the ALLHAT trial (Antihypertensive and Lipid Lowering treatment to prevent Heart Attack Trial) of ~33,000 hypertensive individuals randomised to receive either Lisinopril or Calcium channel blocker, there were small differences in mean BP between treatment groups, but large differences in BPV which paralleled the differences in stroke risk.(ALLHAT Collaborative Research Group 2002.) Stroke risk and BPV were higher in the group treated with Lisinopril, compared to the group treated with calcium channel blockers. (Rothwell et al. 2010b, ALLHAT Collaborative Research Group 2002.) In post hoc analyses of the ASCOT-BPLA trial (Anglo-Scandinavian Cardiac Outcomes Trial- Blood Pressure Lowering Arm), episodic hypertension was greater in the group treated with B blockers, compared to the group treated with calcium channel blockers.(Rothwell 2010b.) Subsequent post-hoc analyses of data from all trials of antihypertensive lowering drugs have shown consistent effects of drug class on group BPV; reduced by calcium channel blockers and increased by B blockers.(Rothwell et al. 2010a.)

Testable future hypotheses would include clinical trials of calcium channel blockers (versus placebo, or other antihypertensive drugs eg ACE-/B blockers) in patients

with normotension and BPV variability versus those with stable normotension. Similar trials could also include populations with treated hypertension, with and without residual BP variability. If these trials demonstrated calcium channel blockers reduced BPV and subsequent stroke risk, it would create new opportunities for stroke prevention, particularly amongst individuals with normal mean BP. It has also been suggested that statins reduce BPV which may explain their additional protective effects beyond cholesterol reduction. (Rothwell 2010b). Finally, the overall effect of calcium channel blockers on BPV reduction has been shown to be modest.(Rothwell 2010b). This means that there is potential in future for development of newer drugs, with potentially greater impact on BPV, which could be used to further reduce BPV related stroke risk.

References

Aboa-Eboule C, Mengue D, Benzenine E, Hommel M, Giroud M, Bejot Y & Quantin C (2013) How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol,* 260**,** 605-13.

Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL & Marsh EE (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke,* 24**,** 35-41.

Adamson J, Beswick A & Ebrahim S (2004) Is stroke the most common cause of disability? *J Stroke Cerebrovasc Dis,* 13**,** 171-7.

Aho K, Harmsen P, Hatano S, Marquardsen J, Smirnov VE & Strasser T (1980) Cerebrovascular disease in the community: results of a WHO collaborative study. *Bull World Health Organ,* 58**,** 113-30.

Albers GW, Caplan LR, Easton JD, Fayad PB, Mohr JP, Saver JL & Sherman DG (2002) Transient ischemic attack--proposal for a new definition. *N Engl J Med,* 347**,** 1713-6.

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. (2002) Major outcomes in high-risk hypertensive patients randomised to angiotensin-converting enzyme inhibitor or calcium channel blocker versus diuretic: The Antihypertensive and Lipid Lowering Treatment to prevent Heart Attack Trial. *JAMA, 288,* 2981-97.

Altman DG & Bland JM (1986) Comparison of methods of measuring blood pressure. *J Epidemiol Community Health,* 40**,** 274-7.

Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA & Hennerici MG (2009a) Classification of Stroke Subtypes. *Cerebrovasc Dis,* 27**,** 493-501.

Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA & Hennerici MG (2009b) New approach to stroke subtyping: the A-S-C-O (phenotypic) classification of stroke. *Cerebrovasc Dis,* 27**,** 502-8.

Andersen K, Olsen KTS, Dehlendorff C & Kammersgaard LP (2009) Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. *Stroke,* 40**,** 2068-72.

Andrade SE, Harrold LR, Tjia J, Cutrona SL, Saczynski JS, Dodd KS, Goldberg RJ & Gurwitz JH (2012) A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf,* 21 Suppl 1**,** 100-28.

Appelros P, Hogeras N & Terent A (2003) Case ascertainment in stroke studies: the risk of selection bias. *Acta Neurol Scand,* 107**,** 145-9.

Appelros P, & Terent A (2011) Validation of the Swedish inpatient and cause of death registers in the context of stroke. *Acta Neurol Scand,* 123**,** 289-93.

Arsava EM, Ballabio E, Benner T, Cole JW, Delgado-Martinez MP, Dichgans M, Fazekas F, Furie KL, Illoh K, Jood K, Kittner S, Lindgren AG, Majersik JJ, Macleod MJ, Meurer WJ, Montaner J, Olugbodi AA, Pasdar A, Redfors P, Schmidt R, Sharma P, Singhal AB, Sorensen AG, Sudlow C, Thijs V, Worrall BB, Rosand J & Ay H (2010) The Causative Classification of Stroke system: an international reliability and optimization study. *Neurology,* 75**,** 1277-84.

Asayama K, Kikuya M, Schutte R, Thijs L, Hosaka M, Satoh M, Hara A, Obara T, Inoue R, Metoki H, Hirose T, Ohkubo T, Staessen JA & Imai Y (2013) Home blood pressure variability as cardiovascular risk factor in the population of Ohasama. *Hypertension,* 61**,** 61-9.

Asdaghi N, Jeerakathil T, Hameed B, Saini M, McCombe JA, Shuaib A, Emery D & Butcher K (2011) Oxfordshire community stroke project classification poorly differentiates small cortical and subcortical infarcts. *Stroke,* 42**,** 2143-8.

Asia Pacific Cohort Studies Collaboration (2005) Joint effects of systolic blood pressure and serum cholesterol on cardiovascular disease in the Asia Pacific region. *Circulation,* 112**,** 3384-90.

Atiya M, Kurth T, Berger K, Buring JE & Kase CS (2003) Interobserver Agreement in the Classification of Stroke in the Women's Health Study. *Stroke,* 34**,** 565-567.

Ay H. (2010) Advances in the diagnosis of etiologic subtypes of ischemic stroke. *Curr Neurol Neurosci Rep,* 10**,** 14-20.

Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, Ayata C, Towfighi A, Smith EE, Chong JY, Koroshetz WJ & Sorensen AG (2007) A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke,* 38**,** 2979-84.

Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG & Koroshetz WJ (2005a) An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol,* 58**,** 688-97.

Ay H, Koroshetz WJ, Benner T, Vangel MG, Wu O, Schwamm LH & Sorensen AG (2005b) Transient ischemic attack with infarction: a unique syndrome? *Ann Neurol,* 57**,** 679-86.

Bamford J, Sandercock P, Dennis M, Burn J & Warlow C (1991) Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet,* 337**,** 1521-6.

Bamford J, Sandercock P, Warlow C & Gray M (1986) Why Are Patients With Acute Stroke Admitted To Hospital? *British Medical Journal (Clinical Research Edition),* 292**,** 1369-1372.

Barer D (1996) Correcting outcome data for case mix in stroke medicine. 313**,** 1005-1006.

Barr EL, Tonkin AM, Welborn TA & Shaw JE (2009) Validity of self-reported cardiovascular disease events in comparison to medical record adjudication and a statewide hospital morbidity database: the AusDiab study. *Internal Medicine Journal,* 39**,** 49-53.

Bejot Y, Mehta Z, Giroud M & Rothwell PM (2013) Impact of completeness of ascertainment of minor stroke on stroke incidence: implications for ideal study methods. *Stroke,* 44**,** 1796-802.

Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, Jackson CA, Traylor M, Strange A, Su Z, Band G, Syme PD, Malik R, Pera J, Norrving B, Lemmens R, Freeman C, Schanz R, James T, Poole D, Murphy L, Segal H, Cortellini L, Cheng YC, Woo D, Nalls MA, Müller-Myhsok B, Meisinger C, Seedorf U, Ross-Adams H, Boonen S, Wloch-Kopec D, Valant V, Slark J, Furie K, Delavaran H, Langford C, Deloukas P, Edkins S, Hunt S, Gray E, Dronov S, Peltonen L, Gretarsdottir S, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Boncoraglio GB, Parati EA, Attia J,

Holliday E, Levi C, Franzosi MG, Goel A, Helgadottir A, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Duncanson A, Jankowski J, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer JR, Trembath RC, Viswanathan AC, Wood NW, Worrall BB, Kittner SJ, Mitchell BD, Kissela B, Meschia JF, Thijs V, Lindgren A, Macleod MJ, Slowik A, Walters M, Rosand J, Sharma P, Farrall M, Sudlow CL, Rothwell PM, Dichgans M, Donnelly P, Markus HS, I. S. G. C. (ISGC) & W. T. C. C. C. (WTCCC2) (2012) Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet,* 44**,** 328-33.

Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP & Matchar DB (1997) Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology,* 49**,** 660-4.

Berger K, Kase CS & Buring JE (1996) Interobserver Agreement in the Classification of Stroke in the Physicians' Health Study. *Stroke,* 27**,** 238-242.

Bergmann MM, Byers T, Freedman DS & Mokdad A (1998) Validity of self-reported diagnoses leading to hospitalization: a comparison of self-reports with hospital records in a prospective study of American adults. *Am J Epidemiol,* 147**,** 969-77.

Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, Jimenez-Conde J, Hansen BM, Fernandez-Cadenas I, Cortellini L, Ayres A, Schwab K, Juchniewicz K, Urbanik A, Rost NS, Viswanathan A, Seifert-Held T, Stoegerer EM, Tomas T, Rabionet R, Estivill X, Brown DL, Silliman SL, Selim M, Worrall BB, Meschia JF, Montaner J, Lindgren A, Roquer J, Schmidt R, Greenberg SM, Slowik A, Broderick JP, Woo D & Rosand J (2010) Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann Neurol,* 68**,** 934-43.

Biobank 2006. Protocol for a large scale prospective epidemiological resource. www.ukbiobank.ac.uk/resources.

Bots ML, Looman SJ, Koudstaal PJ, Hofman A, Hoes AW & Grobbee ED (1996) Prevalence of Stroke in the General Population: The Rotterdam Study. *Stroke,* 27**,** 1499-1501.

Brickman AM, Reitz C, Luchsinger JA, Manly JJ, Schupf N, Muraskin J, DeCarli C, Brown TR & Mayeux R (2010) Long-term blood pressure fluctuation and cerebrovascular disease in an elderly cohort. *Arch Neurol,* 67**,** 564-9.

Britton A, Milne B, Butler T, Sanchez-Galvez A, Shipley M, Rudd A, Wolfe CD, Bhalla A & Brunner EJ (2012) Validating self-reported strokes in a longitudinal UK cohort study (Whitehall II): Extracting information from hospital medical records versus the Hospital Episode Statistics database. *BMC Med Res Methodol,* 12**,** 83.

Brown L, Cai T & DasGupta A (2001) Interval estimation for a binomial proportion. *Statistical Science,* 16**,** 101-133.

Brown MM, Rudd A & McGovern R (2003a) Transient Ischemic Attack — Proposed New Definition. *New England Journal of Medicine,* 348**,** 1607-1609.

Brown MM, Rudd A & McGovern R (2003b) Transient ischemic attack--proposed new definition. *N Engl J Med,* 348**,** 1607-9; author reply 1607-9.

Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P & Faiz OD (2011) Systematic review of discharge coding accuracy. *Journal of Public Health*.
Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J & Elliott P (2009) Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *International Journal of Epidemiology,* 38**,** 263-273.
Campbell SE, Campbell MK, Grimshaw JM & Walker AE (2001) A systematic review of discharge coding accuracy. *J Public Health Med,* 23**,** 205-11.
Carr MJ, Bao Y, Pan J, Cruickshank K & McNamee R (2012) The predictive ability of blood pressure in elderly trial patients. *J Hypertens,* 30**,** 1725-33.
Castle J, Mlynash M, Lee K, Caulfield AF, Wolford C, Kemp S, Hamilton S, Albers GW & Olivot JM (2010) Agreement regarding diagnosis of transient ischemic attack fairly low among stroke-trained neurologists. *Stroke,* 41**,** 1367-70.
Celis H & Fagard RH (2004) White-coat hypertension: a clinical review. *Eur J Intern Med,* 15**,** 348-357.
Ch'Ng A, French D & McLean N (2008) Long term perspectives of stroke support group members. *Journal of Health Psychology,* 13**,** 1136.
Chalela JA, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L & Warach S (2007) Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet,* 369**,** 293-8.
Chapleau MW, Cunningham JT, Sullivan MJ, Wachtel RE, Abboud FM (1995) Structural versus functional modulation of the arterial baroreflex. *Hypertension,* 26, 341-7.
Chen N, Zhou M & Wang Y (2013) Inter-rater reliability of the A-S-C-O classification system for ischemic stroke. *Journal of Clinical Neuroscience,* 20**,** 410-412.
Chen PH, Gao S, Wang YJ, Xu AD, Li YS & Wang D (2012) Classifying Ischemic Stroke, from TOAST to CISS. *CNS Neurosci Ther,* 18**,** 452-6.
Chen Z, Lee L, Chen J, Collins R, Wu F, Guo Y, Linksted P & Peto R (2005) Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol,* 34**,** 1243-9.
Chisholm J. (1990) The Read clinical classification. *BMJ,* 300**,** 1092.
Choi S, Clifton G, Marmarou A & Miller E (2002a) Misclassification and treatment effect on primary outcome measures in clinical trials of severe neurotrauma. *Journal of Neurotrauma,* 19**,** 17-22.
Chowdhury EK, Owen A, Krum H, Wing LM, Nelson MR, Reid CM & Second Australian National Blood Pressure Study Management Committee (2014) Systolic blood pressure variability is an important predictor of cardiovascular outcomes in elderly hypertensive patients. *J Hypertens,* 32**,** 525-33.
Clarke, R., Shipley M, Lewington S, Youngman L, Collins R, Marmot M & Peto R (1999) Underestimation of Risk Associations Due to Regression Dilution in Long-term Follow-up of Prospective Studies. *American Journal of Epidemiology,* 150, 341-353.
Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, Hennekens CH & Speizer FE (1986) Validation of questionnaire information on risk

factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol,* 123**,** 894-900.
Collins R. (2012) What makes UK Biobank special? *Lancet,* 379**,** 1173-4.
Cotter PE, Belham M & Martin PJ (2012) Towards understanding the cause of stroke in young adults utilising a new stroke classification system (A-S-C-O). *Cerebrovasc Dis,* 33**,** 123-7.
D'Olhaberriague L, Litvan I, Mitsias P & Mansbach HH (1996) A Reappraisal of Reliability and Validity Studies in Stroke. *Stroke,* 27**,** 2331-2336.
Davenport RJ, Dennis MS & Warlow CP (1996) The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. *Health Bull (Edinb),* 54**,** 402-5.
Davé S & Petersen I (2009) Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiology and Drug Safety,* 18**,** 704-707.
Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, Kivimaki M, Timmis AD, Smeeth L & Hemingway H (2012) Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *International Journal of Epidemiology,* 41**,** 1625-1638.
Derby CA, Lapane KL, Feldman HA & Carleton RA (2000) Trends in Validated Cases of Fatal and Nonfatal Stroke, Stroke Classification, and Risk Factors in Southeastern New England, 1980 to 1991: Data From the Pawtucket Heart Health Program. *Stroke,* 31**,** 875-881.
Derby CA, Lapane KL, Feldman HA & Carleton RA (2001) Possible Effect of DRGs on the Classification of Stroke: Implications for Epidemiological Surveillance. *Stroke,* 32**,** 1487-1491.
deVet HCW, Riphagen EA, Aertgeerts B, Pewsner D (2008) Chapter 7: Searching for Studies. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy,* version 0.4**,** available: www.cochrane-handbook.org.
deVet HCW, Terwee CB, Knol DL & Bouter LM (2006) When to use agreement versus reliability measures. *J Clin Epidemiol,* 59**,** 1033-9.
Dewey H, Macdonell R, Donnan G, McNeil J, Freeman E, Thrift A & Sharples C (2001) Inter-rater reliability of stroke sub-type classification by neurologists and nurses within a community-based stroke incidence study. *Journal of Clinical Neuroscience,* 8**,** 14-17.
Diaz KM, Tanner RM, Falzon L, Levitan EB, Reynolds K, Shimbo D & Muntner P (2014) Visit-to-visit variability of blood pressure and cardiovascular disease and all-cause mortality: a systematic review and meta-analysis. *Hypertension,* 64**,** 965-82.
Dixon J, Sanderson C, Elliott P, Walls P, Jones J & Petticrew M (1998) Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *Journal of Public Health,* 20**,** 63-69.
Dolan E, Stanton A, Thijs L, Hinedi K, Atkins N, McClory S, Den Hond E, McCormack P, Staessen JE & O'Brien E (2005) Superiority of ambulatory over clinic blood pressure measurement in predicting mortality: the Dublin outcome study. *Hypertension,* 46**,** 156-61.

Doubal FN, MacGillivray TJ, Hokke PE, Dhillon B, Dennis MS & Wardlaw JM (2009) Differences in retinal vessels support a distinct vasculopathy causing lacunar stroke. *Neurology,* 72**,** 1773-8.

Drizd T, Dannerberg A & Engel A (1986) Blood pressure levels in persons 18-74 years of age in 1976-80, and trends in blood pressure from 1960 to 1980 in the United States. *Vital and Health Statistics,* 234, 1-68

Eguchi K, Hoshide S, Schwartz JE, Shimada K & Kario K (2012) Visit-to-visit and ambulatory blood pressure variability as predictors of incident cardiovascular events in patients with hypertension. *Am J Hypertens,* 25**,** 962-8.

Ellekjær H, Holmen J, Krüger Ø & Terent A (1999) Identification of Incident Stroke in Norway: Hospital Discharge Data Compared With a Population-Based Stroke Register. *Stroke,* 30**,** 56-60.

Englert H, Muller-Nordhorn J, Seewald S, Sontag F, Voller H, Meyer-Sabelleck W, Wegscheider K, Windler E, Katus H, Willich S (2010) Is patient self-report and adequate tool for monitoring cardiovascular conditions in patients with hypercholesterolaemia? *J Public Health,* 32, 387-394.

Engstad T, Bonaa KH, Viitanen M. (2000) Validity of self-reported stroke. The Tromso study. *Stroke,* 31**,** 1602-1607.

Eto M, Toba K, Akishita M, Kozaki K, Watanabe T, Kim S, Hashimoto M, Ako J, Iijima K, Sudoh N, Yoshizumi M & Ouchi Y (2005) Impact of blood pressure variability on cardiovascular events in elderly patients with hypertension. *Hypertens Res,* 28**,** 1-7.

Falaschetti E, Chaudhury M, Mindell J & Poulter N (2009) Continued Improvement in Hypertension Management in England: Results From the Health Survey for England 2006. *Hypertension,* 53**,** 480-486.

Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, Moran AE, Sacco RL, Anderson L, Truelsen T, O'Donnell M, Venketasubramanian N, Barker-Collo S, Lawes CM, Wang W, Shinohara Y, Witt E, Ezzati M, Naghavi M Murray C, Global Burden of Diseases, Injuries and Risk Factors Study (GBD 2010) & GBD Stroke Experts Group (2014) Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet,* 383**,** 245-54.

Ferro JM, Falcão I, Rodrigues G, Canhão P, Melo TP, Oliveira V, Pinto AN, Crespo M & Salgado AV (1996) Diagnosis of transient ischemic attack by the nonneurologist. A validation study. *Stroke,* 27**,** 2225-9.

Ferro JM, Pinto AN, Falcão I, Rodrigues G, Ferreira J, Falcão F, Azevedo E, Canhão P, Melo TP, Rosas MJ, Oliveira V & Salgado AV (1998) Diagnosis of stroke by the nonneurologist. A validation study. *Stroke,* 29**,** 1106-9.

Fiebach JB, Schellinger PD, Jansen O, Meyer M, Wilde P, Bender J, Schramm P, Juttler E, Oehler J, Hartmann M, Hahnel S, Knauth M, Hacke W & Sartor K (2002) CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. *Stroke,* 33**,** 2206-10.

Floras JS (2013) Blood Pressure Variability: A Novel and Important Risk Factor. *Canadian Journal of Cardiology,* 29**,** 557-563.

Flynn RW, Macdonald TM, Schembri N, Murray GD & Doney AS (2010) Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. *Pharmacoepidemiol Drug Saf,* 19**,** 843-7.

Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, Brooks CJ, Thompson S, Bodger O, Couch T & Leake K (2009) The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research,* 9**,** 157-157.

Fourrier-Reglat A, Cuong HM, Lassalle R, Depont F, Robinson P, Droz-Perroteau C, Pariente A, Begaud B, Blin P & Moore N (2010) Concordance between prescriber- and patient-reported previous medical history and NSAID indication in the CADEUS cohort. *Pharmacoepidemiol Drug Saf,* 19**,** 474-81.

Frattola A, Parati G, Cuspidi C, Albini F & Mancia G (1993) Prognostic value of 24-hour blood pressure variability. *J Hypertens,* 11**,** 1133-7.

Fure B, Wyller TB & Thommessen B (2005) TOAST criteria applied in acute ischemic stroke. *Acta Neurol Scand,* 112**,** 254-8.

Gaist D, Wallander MA, Gonzalez-Perez A & Garcia-Rodriguez LA (2013) Incidence of hemorrhagic stroke in the general population: validation of data from The Health Improvement Network. *Pharmacoepidemiol Drug Saf,* 22**,** 176-82.

Gao S, Hendrie HC, Wang C, Stump TE, Stewart JC, Kesterson J, Clark DO & Callahan CM (2014) Redefined blood pressure variability measure and its association with mortality in elderly primary care patients. *Hypertension,* 64**,** 45-52.

Gao S, Wang J, Xu AD, Li YS & Wang DZ (2011) Chinese ischemic stroke subclassification. *Frontiers in Neurology,* 2**,** 1-5.

Ghandehari K, Mouradian M, Izadi Z & Salam A (2005) Reliability of Practical Iranian Criteria (PIC) for classification of brain infarct. *Archives of Iranian Medicine,* 8**,** 96-99.

Giroud M, Lemesle M, Quantin C, Vourch M, Becker F, Milan C, Brunet-Lecomte P & Dumas R (1997) A hospital-based and a population-based stroke registry yield different results: the experience in Dijon, France. *Neuroepidemiology,* 16**,** 15-21.

Goldstein LB (1998) Accuracy of ICD-9-CM Coding for the Identification of Patients With Acute Ischemic Stroke: Effect of Modifier Codes. *Stroke,* 29**,** 1602-1604.

Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chilukuri V, Armstrong SB & Horner RD (2001) Improving the Reliability of Stroke Subgroup Classification Using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Criteria. *Stroke,* 32**,** 1091-1097.

Gordon DL, Bendixen BH, Adams HP, Clarke W, Kappelle LJ & Woolson RF (1993) Interphysician agreement in the diagnosis of subtypes of acute ischemic stroke: implications for clinical trials. The TOAST Investigators. *Neurology,* 43**,** 1021-7.

Grimes DA & Schulz KF (2002) Cohort studies: marching towards outcomes. *The Lancet,* 359**,** 341-345.

Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, Meitinger T, Wichmann E, Mitchell BD, Furie K, Slowik A, Rich SS, Syme PD, MacLeod MJ, Meschia JF, Rosand J, Kittner SJ, Markus HS, Müller-

Myhsok B, Dichgans M & International Stroke Genetics Consortium (2009) Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol,* 65**,** 531-9.

Gulliford MC, Charlton J, Ashworth M, Rudd AG & Toschke AM for the eCRT Research Team (2009) Selection of Medical Diagnostic Codes for Analysis of Electronic Patient Records. Application to Stroke in a Primary Care Database. *PLoS ONE,* 4**,** e7168.

Haesebaert J, Termoz A, Polazzi S, Mouchoux C, Mechtouff L, Derex L, Nighoghossian N & Schott AM (2013) Can Hospital Discharge Databases Be Used to Follow Ischemic Stroke Incidence? *Stroke,* 44**,** 1770-1774.

Hajat C, Coshall C, Rudd AG, Patel M, Bhalla A, Howard R & Wolfe CD (2001) The inter- and intraobserver reliabilities of a new classification system for ischaemic stroke: the South London Stroke Register. *Journal of the Neurological Sciences,* 190**,** 79-85.

Han SW, Kim SH, Lee JY, Chu CK, Yang JH, Shin HY, Nam HS, Lee BI & Heo JH (2007) A new subtype classification of ischemic stroke based on treatment and etiologic mechanism. *Eur Neurol,* 57**,** 96-102.

Hand PJ, Haisma JA, Kwan J, Lindley RI, Lamont B, Dennis MS & Wardlaw JM (2006a) Interobserver Agreement for the Bedside Clinical Assessment of Suspected Stroke. *Stroke,* 37**,** 776-780.

Hand P J, Kwan J, Lindley RI, Dennis MS & Wardlaw JM (2006b) Distinguishing Between Stroke and Mimic at the Bedside: The Brain Attack Study. *Stroke,* 37**,** 769-775.

Harlow SD & Linet MS (1989) Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol,* 129**,** 233-48.

Harriss LR, Ajani AE, Hunt D, Shaw J, Chambers B, Dewey H, Frayne J, Beauchamp A, Duvé K, Giles GG, Harrap S, Magliano DJ, Liew D, McNeil J, Peeters A, Stebbing M, Wolfe R & Tonkin A (2011) Accuracy of national mortality codes in identifying adjudicated cardiovascular deaths. *Australian and New Zealand Journal of Public Health,* 35**,** 466-476.

Hasan M, Meara RJ & Bhowmick BK. 1995. The Quality of Diagnostic Coding in Cerebrovascular Disease. *Int J Qual Health care,* 7, 407-410

Hashimoto T, Kikuya M, Ohkubo T, Satoh M, Metoki H, Inoue R, Asayama K, Kanno A, Obara T, Hirose T, Hara A, Hoshi H, Totsune K, Satoh H, Sato H & Imai Y (2012) Home blood pressure level, blood pressure variability, smoking, and stroke risk in Japanese men: the Ohasama study. *Am J Hypertens,* 25**,** 883-91.

Hastie CE, Jeemon P, Coleman H, McCallum L, Patel R, Dawson J, Sloan W, Meredith P, Jones GC, Muir S, Walters M, Dominiczak AF, Morrison D, McInnes GT & Padmanabhan S (2013) Long-term and ultra long-term blood pressure variability during follow-up and mortality in 14,522 patients with hypertension. *Hypertension,* 62**,** 698-705.

Hata J, Arima H, Rothwell PM, Woodward M, Zoungas S, Anderson C, Patel A, Neal B, Glasziou P, Hamet P, Mancia G, Poulter N, Williams B, Macmahon S, Chalmers J & ADVANCE Collaborative Group (2013) Effects of visit-to-visit variability in systolic blood pressure on macrovascular and microvascular complications in patients with type 2 diabetes mellitus: the ADVANCE trial. *Circulation,* 128**,** 1325-34.

Hata Y, Kimura Y, Muratani H, Fukiyama K, Kawano Y, Ashida T, Yokouchi M, Imai Y, Ozawa T, Fujii J & Omae T (2000) Office blood pressure variability as a predictor of brain infarction in elderly hypertensive patients. *Hypertens Res,* 23**,** 553-60.

Hatano S (1976) Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ,* 54**,** 541-53.

Health and Social Care Information Centre 2012. Prescriptions dispensed in the community: England, statistics for 2001 to 2011.

Heckbert SR, Kooperberg C, Safford MM, Psaty BM, Hsia J, McTiernan A, Gaziano JM, Frishman WH & Curb JD (2004) Comparison of self-report, hospital discharge codes, and adjudication of cardiovascular events in the Women's Health Initiative. *Am J Epidemiol,* 160**,** 1152-8.

Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, Timmis A & Hemingway H (2013) Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ,* 346**,** f2350.

Herrett E, Thomas SL, Schoonen WM, Smeeth L & Hall AJ (2010) Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British Journal of Clinical Pharmacology,* 69**,** 4-14.

Heuschmann PU, Kolominsky-Rabas PL, Weber M, Boehner J, Erbguth F & Neundoerfer B (1999) Feasibility and inter-rater reliability of OCSP and TOAST classification during the acute stage of stroke. *Cerebrovasc Dis,* 9**,** 1-128.

Higgins JGS (2011) Cochrane Handbook for systematic reviews of interventions. *The Cochrane Collaboration,* version 5.1.0.**,** available: www.cochrane-handbook.org.

Hocht C (2013) Blood pressure variability: prognostic value and therapeutic implications. *ISRN Hypertension*. Article ID 398485

Hodgkinson J, Mant J, Martin U, Guo B, Hobbs FD, Deeks JJ, Heneghan C, Roberts N & McManus RJ (2011) Relative effectiveness of clinic and home blood pressure monitoring compared with ambulatory blood pressure monitoring in diagnosis of hypertension: systematic review. *BMJ,* 342**,** d3621.

Hodgkinson JA, Tucker KL, Crawford C, Greenfield SM, Heneghan C, Hinton L, Khan K, Locock L, Mackillop L, McCourt C, Selwood M & McManus RJ (2014) Is self monitoring of blood pressure in pregnancy safe and effective? *BMJ,* 349**,** g6616.

Howard SC & Rothwell PM (2009) Reproducibility of measures of visit-to-visit variability in blood pressure after transient ischaemic attack or minor stroke. *Cerebrovasc Dis,* 28**,** 331-40.

Hsieh YT, Tu ST, Cho TJ, Chang SJ, Chen JF & Hsieh MC (2012) Visit-to-visit variability in blood pressure strongly predicts all-cause mortality in patients with type 2 diabetes: a 5·5-year prospective analysis. *Eur J Clin Invest,* 42**,** 245-53.

http://apps.who.int/classifications/icd10/browse/2010/en.

http://biobank.ctsu.ox.ac.uk/crystal

http://systems.hscic.gov.uk/data/uktc/readcodes.

http://systems.hscic.gov.uk/data/uktc/readcodes/drugandappliancedictionary.

http://www.caliberresearch.org/.

http://www.cdc.gov/nchs/icd.htm.
http://www.cprd.com/Bibliography/Researchpapers.asp.
http://www.icd9data.com/.
http://www.isdscotland.org/.
http://www.nice.org.uk/guidance/cg127/chapter/1-recommendations.
http://www.saildatabank.com/.
http://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/publications.
http://www.ukbiobank.ac.uk/
http://www.wales.nhs.uk/.
http://www.wales.nhs.uk/sites3/page.cfm?orgid=480&pid=10486.
https://isd.hscic.gov.uk/trud3/user/guest/group/0/home.

Ives DG, Fitzpatrick AL, Bild DE, Psaty BM, Kuller LH, Crowley PM, Cruise RG & Theroux S (1995) Surveillance and ascertainment of cardiovascular events: The Cardiovascular Health Study. *Annals of Epidemiology,* 5**,** 278-285.

Jin J, Sklar GE, Oh VMS & Li SC (2008) Factors affecting therapeutic compliance: a review from the patient's perspective. *Theraputics and Clinical Risk Management,* 4**,** 269-286.

Jackson C. & Sudlow C (2005) Are lacunar strokes really different? A systematic review of differences in risk factor profiles between lacunar and nonlacunar infarcts. *Stroke,* 36**,** 891-901.

Jackson CA, Hutchison A, Dennis MS, Wardlaw JM, Lewis SC & Sudlow CL (2009) Differences between ischemic stroke subtypes in vascular outcomes support a distinct lacunar ischemic stroke arteriopathy: a prospective, hospital-based study. *Stroke,* 40**,** 3679-84.

Jackson CA, Hutchison A, Dennis MS, Wardlaw JM, Lindgren A, Norrving B, Anderson CS, Hankey GJ, Jamrozik K, Appelros P & Sudlow CLM (2010) Differing Risk Factor Profiles of Ischemic Stroke Subtypes: Evidence for a Distinct Lacunar Arteriopathy? *Stroke,* 41**,** 624-629.

Jackson CA & Sudlow CL (2006) Is hypertension a more frequent risk factor for deep than for lobar supratentorial intracerebral haemorrhage? *J Neurol Neurosurg Psychiatry,* 77**,** 1244-52.

Jaffar S, Leach A, Smith P, Cutts F & Greenwood G (2003a) Effects of misclassification of causes of death on the power of a trial to assess the efficacy of a pneumococcal conjugate vaccine in The Gambia. *International Journal of Epidemiology,* 32**,** 430-436.

Jin YP, Di Legge S, Østbye T, Feightner JW, Saposnik G & Hachinski V (2010) Is Stroke History Reliably Reported by Elderly with Cognitive Impairment? A Community-Based Study. *Neuroepidemiology,* 35**,** 215-220.

Johansson JK, Niiranen TJ, Puukka PJ & Jula AM (2012) Prognostic value of the variability in home-measured blood pressure and heart rate: the Finn-Home Study. *Hypertension,* 59**,** 212-8.

Johnsen SP, Overvad K, Sørensen HT, Tjønneland A & Husted SE (2002) Predictive value of stroke and transient ischemic attack discharge diagnoses in The Danish National Registry of Patients. *Journal of Clinical Epidemiology,* 55**,** 602-607.

Johnson CJ, Kittner SJ, McCarter RJ, Sloan MA, Stern BJ, Buchholz D & Price TR (1995) Interrater Reliability of an Etiologic Classification of Ischemic Stroke. *Stroke,* 26**,** 46-51.

Jordan K, Porcheret M & Croft P (2004) Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam Pract,* 21**,** 396-412.
Kessler C, Freyberger H, Dittman V & Ringelstein E (1991) Interrater reliability in the assessment of neurovascular diseases. *Cerebrovasc Dis,* 1**,** 43-48.
Khan NF, Harrison SE & Rose PW (2010) Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract,* 60**,** e128-36.
Kikuya M, Ohkubo T, Metoki H, Asayama K, Hara A, Obara T, Inoue R, Hoshi H, Hashimoto J, Totsune K, Satoh H & Imai Y (2008) Day-by-day variability of blood pressure and heart rate at home as a novel predictor of prognosis: the Ohasama study. *Hypertension,* 52**,** 1045-50.
Kimberlin CL & Winterstein AG (2008) Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm,* 65**,** 2276-84.
Kirkman MA, Mahattanakul W, Gregson BA & Mendelow AD (2009) The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurol Belg,* 109**,** 114-9.
Klatsky AL, Friedman GD, Sidney S, Kipp H, Kubo A & Armstrong MA (2005) Risk of hemorrhagic stroke in Asian American ethnic groups. *Neuroepidemiology,* 25**,** 26-31.
Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B & Heuschmann PU (2001) Epidemiology of ischemic stroke subtypes according to TOAST criteria: incidence, recurrence, and long-term survival in ischemic stroke subtypes: a population-based study. *Stroke,* 32**,** 2735-40.
Koster M, Asplund K, Johansson A & Stegmayr B (2013a) Refinement of Swedish Administrative Registers to monitor stroke events on the national level. . *Neuroepidemiology,* 40**,** 240-246.
Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M & Streiner DL (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol,* 64**,** 96-106.
Krarup LH, Boysen G, Janjua H, Prescott E & Truelsen T (2007) Validity of stroke diagnoses in a National Register of Patients. *Neuroepidemiology,* 28**,** 150-4.
Kriegsman DMW, Penninx BWJH, Van Eijk JK, Boeke AJP & Deeg DJH (1996) Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly: A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *Journal of Clinical Epidemiology,* 49**,** 1407-1417.
Lakshminarayan K, Anderson DC, Jacobs DR, Barber CA & Luepker RV (2009) Stroke rates: 1980-2000: the Minnesota Stroke Survey. *Am J Epidemiol,* 169**,** 1070-8.
Lawes CM, Vander Hoorn S & Rodgers A (2008) Global burden of blood-pressure-related disease, 2001. *Lancet,* 371**,** 1513-8.
Leibson CL, Naessens JM, Brown RD & Whisnant JP (1994) Accuracy of hospital discharge abstracts for identifying stroke. *Stroke,* 25**,** 2348-55.
Leone MA, Capponi A, Varrasi C, Tarletti R & Monaco F (2004) Accuracy of the ICD-9 codes for identifying TIA and stroke in an Italian automated database. *Neurological Sciences,* 25**,** 281-288.

Levitan EB, Kaciroti N, Oparil S, Julius S & Muntner P (2012) Blood pressure measurement device, number and timing of visits, and intra-individual visit-to-visit variability of blood pressure. *J Clin Hypertens (Greenwich),* 14**,** 744-50.

Lewington S, Clarke R, Qizilbash N, Peto R & Collins R (2002) Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet,* 360**,** 1903-13.

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, Amann M, Anderson HR, Andrews KG, Aryee M, Atkinson C, Bacchus LJ, Bahalim AN, Balakrishnan K, Balmes J, Barker-Collo S, Baxter A, Bell ML, Blore JD, Blyth F, Bonner C, Borges G, Bourne R, Boussinesq M, Brauer M, Brooks P, Bruce NG, Brunekreef B, Bryan-Hancock C, Bucello C, Buchbinder R, Bull F, Burnett RT, Byers TE, Calabria EB, Carapetis J, Carnahan E, Chafe Z, Charlson F, Chen H, Chen JS, Cheng AT, Child JC, Cohen A, Colson KE, Cowie BC, Darby S, Darling S, Davis A, Degenhardt L, Dentener F, Des Jarlais DC, Devries K, Dherani M, Ding EL, Dorsey ER, Driscoll T, Edmond K, Ali SE, Engell RE, Erwin PJ, Fahimi S, Falder G, Farzadfar F, Ferrari A, Finucane MM, Flaxman S, Fowkes FG, Freedman G, Freeman MK, Gakidou E, Ghosh S, Giovannucci E, Gmel G, Graham K, Grainger R, Grant B, Gunnell D, Gutierrez HR, Hall W, Hoek HW, Hogan A, Hosgood HD, Hoy D, Hu H, Hubbell BJ, Hutchings SJ, Ibeanusi SE, Jacklyn GL, Jasrasaria R, Jonas JB, Kan H, Kanis JA, Kassebaum N, Kawakami N, Khang YH, Khatibzadeh S, Khoo JP, Kok C, Laden F et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2224-60.

Lindley RI, Wang JJ, Wong MC, Mitchell P, Liew G, Hand P, Wardlaw J, De Silva DA, Baker M, Rochtchina E, Chen C, Hankey GJ, Chang HM, Fung VS, Gomes L, Wong TY & Multi-Centre Retina and Stroke Study (MRCS) Group (2009) Retinal microvasculature in acute lacunar stroke: a cross-sectional study. *Lancet Neurol,* 8**,** 628-34.

Lindley RI, Warlow CP, Wardlaw JM, Dennis MS, Slattery J & Sandercock PA (1993) Interobserver reliability of a clinical classification of acute cerebral infarction. *Stroke,* 24**,** 1801-4.

Lindsted KE, Fraser GE, Steinkohl M, Beeson WL (1996) Healthy volunteer effect in a cohort study: Temporal resolution in the adventist health study. *J Clin Epidemiol*, 49, 783-90.

Liu L, Reeder B, Shuaib A & Mazagri R (1999) Validity of Stroke Diagnosis on Hospital Discharge Records in Saskatchewan, Canada: Implications for Stroke Surveillance. *Cerebrovascular Diseases,* 9**,** 224-230.

Lopez AD & Mathers CD (2006) Measuring the global burden of disease and epidemiological transitions: 2002-2030. *Ann Trop Med Parasitol,* 100**,** 481-99.

Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo LS, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck

G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne DS, Dorsey SR, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, et al. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2095-128.

Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, Brown G & Leake K (2009) The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak,* 9**,** 3.

MacDonald T, Morant S & Mozzafari E (2007) Drug treatment discontinuation and achievement of target blood pressure and cholesterol in United Kingdom Primary Care. *Current Medical Research and Opinion,* 23**,** 2765-2774.

Machon M, Arriola L, Larranaga N, Amiano P, Moreno-Iribas C, Agudo A, Ardanaz E, Barricarte A, Buckland G, Chirlaque MD, Gavrila D, Huerta JM, Martinez C, Molina E, Navarro C, Quiros JR, Rodriguez L, Sanchez MJ, Gonzalez CA & Dorronsoro M (2013) Validity of self-reported prevalent cases of stroke and acute myocardial infarction in the Spanish cohort of the EPIC study. *J Epidemiol Community Health,* 67**,** 71-5.

MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A & Stamler J (1990) Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet,* 335**,** 765-74.

Magnusson RS (2009) Rethinking global health challenges: towards a 'global compact' for reducing the burden of chronic disease. *Public Health,* 123**,** 265-74.

Mancia G (2012) Short- and long-term blood pressure variability: present and future. *Hypertension,* 60**,** 512-7.

Mancia G, Facchetti R, Parati G & Zanchetti A (2012) Visit-to-visit blood pressure variability in the European Lacidipine Study on Atherosclerosis: methodological aspects and effects of antihypertensive treatment. *J Hypertens,* 30**,** 1241-51.

Mancia G, Messerli F, Bakris G, Zhou Q, Champion A & Pepine CJ (2007) Blood pressure control and improved cardiovascular outcomes in the International Verapamil SR-Trandolapril Study. *Hypertension,* 50**,** 299-305.

Mann S, Craig MWM & Raftery EB (1985) Superiority of 24-hour measurement of blood pressure over clinic values in determining prognosis in hypertension. *Clin Exp Hypertens A,* 7**,** 279-81.

Manolio TA, Bailey-Wilson JE & Collins FS (2006) Genes, environment and the value of prospective cohort studies. *Nat Rev Genet,* 7**,** 812-820.

Mant J, Mant F & Winner S (1997) How good is routine information? Validation of coding for acute stroke in Oxford hospitals. *Health Trends,* 29.

Mant J, McManus R, Hare R & Mayer P (2003) Identification of stroke in the community: a comparison of three methods. *Br J Gen Pract,* 53**,** 520-524.

Marnane M, Duggan CA & Sheehan OC (2009) Improving sub-type classification in TOAST - defined cryptogenic stroke in the North Dublin population Stroke Study. *Cerebrovasc Dis,* 27**,** 101.

Marnane M, Duggan CA, Sheehan OC, Merwick A, Hannon H, Curtin D, Harris D, Williams EB, Horgan G, Kyne L, McCormack PM, Duggan J, Moore A, Crispino-O'Connell G & Kelly PJ (2010) Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, A-S-C-O, and causative classification system: direct comparison in the North Dublin population stroke study. *Stroke,* 41**,** 1579-86.

Masson P, Kelly PJ, Craig JC, Lindley RI & Webster AC (2015) Risk of Stroke in Patients with ESRD. *Clin J Am Soc Nephrol,* 10**,** 1585-92.

Mayo N, Danys I, Carlton J & Scott S (1993) Accuracy of hospital discharge coding for stroke. *Can J Cardiol,* 9**,** 121D-123D.

McArdle PF, Kittner SJ, Ay H, Brown RD, Meschia JF, Rundek T, Wassertheil-Smoller S, Woo D, Andsberg G, Biffi A, Brenner DA, Cole JW, Corriveau R, de Bakker PI, Delavaran H, Dichgans M, Grewal RP, Gwinn K, Huq M, Jern C, Jimenez-Conde J, Jood K, Kaplan RC, Katschnig P, Katsnelson M, Labovitz DL, Lemmens R, Li L, Lindgren A, Markus HS, Peddareddygari LR, Pedersen A, Pera J, Redfors P, Roquer J, Rosand J, Rost NS, Rothwell PM, Sacco RL, Sharma P, Slowik A, Sudlow C, Thijs V, Tiedt S, Valenti R & Worrall BB (2014) Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. *Neurology,* 83**,** 1653-60.

McCormick N, Lacaille D, Bhole V & Avina-Zubieta JA (2014) Validity of Myocardial Infarction Diagnoses in Administrative Databases: A Systematic Review. *PLoS ONE,* 9**,** e92286.

McManus RJ, Wood S, Bray EP, Glasziou P, Hayen A, Heneghan C, Mant J, Padfield P, Potter JF & Hobbs FD (2014) Self-monitoring in hypertension: a web-based survey of primary care physicians. *J Hum Hypertens,* 28**,** 123-7.

Mead GE, Lewis SC, Wardlaw JM, Dennis MS & Warlow CP (2000) How well does the Oxfordshire community stroke project classification predict the site and size of the infarct on brain imaging? *J Neurol Neurosurg Psychiatry,* 68**,** 558-62.

Mena L, Pintos S, Queipo NV, Aizpúrua JA, Maestre G & Sulbarán T (2005) A reliable index for the prognostic significance of blood pressure variability. *J Hypertens,* 23**,** 505-11.

Meredith PA, Perloff D, Mancia G & Pickering T (1995) Blood pressure variability and its implications for antihypertensive therapy. *Blood Press,* 4**,** 5-11.

Meschia J.F, Barrett KM, Chukwudelunzu F, Brown WM, Case LD, Kissela BM, Brown RD, Brott TG, Olson TS, Rich SS, Silliman S & Worrall BB (2006) Interobserver Agreement in the Trial of Org 10172 in Acute Stroke Treatment Classification of Stroke Based on Retrospective Medical Record Review. *Journal of Stroke and Cerebrovascular Diseases,* 15**,** 266-272.

Morrison V, Holmes V, Parveen S, Plumpton C, Clyne W, De Geest S, Dobbels F, Vrijens B, Kardas P & Hughes DA (2015) Predictors of self-reported adherence to antihypertensive medicines: a multi-national cross-sectional survey. *Value Health,* 18**,** 201-216.

Mukherjee D & Patil CG (2011) Epidemiology and the global burden of stroke. *World Neurosurg,* 76**,** S85-90.

Munger MA, Van Tassell BW & LaFleur J (2007) Medication nonadherence: an unrecognized cardiovascular risk factor. *MedGenMed,* 9**,** 58.

Muntner P, Joyce C, Levitan EB, Holt E, Shimbo D, Webber LS, Oparil S, Re R & Krousel-Wood M (2011a) Reproducibility of visit-to-visit variability of blood pressure measured as part of routine clinical care. *J Hypertens,* 29**,** 2332-8.

Muntner P, Shimbo D, Tonelli M, Reynolds K, Arnett DK & Oparil S (2011b) The relationship between visit-to-visit variability in systolic blood pressure and all-cause mortality in the general population: findings from NHANES III, 1988 to 1994. *Hypertension,* 57**,** 160-6.

Murray CJ, Richards MA, Newton JN, Fenton KA, Anderson HR, Atkinson C, Bennett D, Bernabé E, Blencowe H, Bourne R, Braithwaite T, Brayne C, Bruce NG, Brugha TS, Burney P, Dherani M, Dolk H, Edmond K, Ezzati M, Flaxman AD, Fleming TD, Freedman GD, Gunnell D, Hay RJ, Hutchings SJ, Ohno SL, Lozano R, Lyons RA, Marcenes W, Naghavi M, Newton CR, Pearce N, Pope D, Rushton L, Salomon JA, Shibuya K, Vos T, Wang H, Williams HC, Woolf AD, Lopez AD & Davis A (2013) UK health performance: findings of the Global Burden of Disease Study 2010. *Lancet,* 381**,** 997-1020.

Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, L. Baddour M, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne DS, et al. (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2197-223.

Murray J, Young J & Forster A (2007) Review of longer-term problems after disabling stroke. *Reviews in Clinical Gerontology,* 17**,** 277-292.

Myint PK, Sinha S, Luben RN, Bingham SA, Wareham NJ & Khaw KT (2008) Risk factors for first-ever stroke in the EPIC-Norfolk prospective population-based study. *Eur J Cardiovasc Prev Rehabil,* 15**,** 663-9.

Nadarajan V, Perry RJ, Johnson J & Werring DJ (2014) Transient ischaemic attacks: mimics and chameleons. *Practical Neurology,* 14**,** 23-31.

Nagai M. & Kario K (2013) Visit-to-visit blood pressure variability, silent cerebral injury, and risk of stroke. *Am J Hypertens,* 26**,** 1369-76.

Nakano C, Morimoto S, Nakahigashi M, Kusabe M, Ueda H, Someya K, Ichihara A, Iwasaka T & Shiojima I (2015) The relationships between visit-to-visit blood pressure variability and renal and endothelial function in chronic kidney disease. *Hypertens Res,* 38**,** 193-8.

Nam HS, Cha MJ, Kim YD, Kim EH, Park E, Lee HS, Nam CM & Heo JH (2012) Use of a handheld, computerized device as a decision support tool for stroke classification. *European Journal of Neurology,* 19**,** 426-430.

Nicodème R, Albessard A, Amar J, Chamontin B & Lang T (2009) Poor blood pressure control in general practice: in search of explanations. *Arch Cardiovasc Dis,* 102**,** 477-83.

O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, Rangarajan S, Islam S, Pais P, McQueen MJ, Mondo C, Damasceno A, Lopez-Jaramillo P, Hankey GJ, Dans AL, Yusoff K, Truelsen T, Diener HC, Sacco RL, Ryglewicz D, Czlonkowska A, Weimar C, Wang X & Yusuf S (2010) Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet,* 376**,** 112-23.

O'Mahony PG, Dobson R, Rodgers H, James OF & Thomson RG (1995) Validation of a population screening questionnaire to assess prevalence of stroke. *Stroke,* 26**,** 1334-7.

Ohkubo T, Asayama K, Kikuya M, Metoki H, Obara T, Saito S, Hoshi H, Hashimoto J, Totsune K, Satoh H & Imai Y (2004) Prediction of ischaemic and haemorrhagic stroke by self-measured blood pressure at home: the Ohasama study. *Blood Press Monit,* 9**,** 315-20.

Okura Y, Urban LH, Mahoney DW, Jacobsen SJ & Rodeheffer RJ (2004) Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol,* 57**,** 1096-103.

Padfield PL (2010) The case for home monitoring in hypertension. *BMC Med,* 8**,** 55.

Paganini-Hill A & Chao A (1993) Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. *Am J Epidemiol,* 138**,** 101-6.

Palmieri L, Barchielli A, Cesana G, deCampora E, Goldoni C, Sporalore P, Uguccioni M, Vancheri F, Vanuzzo D, Ciccarelli P & Giampaoli S (2007a) The Italian Register of Cardiovascular Diseases: Attack rates and case fatality for cerebrovascular events. *Cerebrovasc Dis,* 24**,** 530-539.

Panayiotou B, Fotherby M, Potter J & Castleden C (1993) The accuracy of diagnostic coding of cerebrovascular disease. *Medical Audit News,* 3**,** 153-155.

Parati G, Ochoa JE, Lombardi C & Bilo G (2013) Assessment and management of blood-pressure variability. *Nat Rev Cardiol,* 10**,** 143-55.

Pierdomenico SD, Di Nicola M, Esposito AL, Di Mascio R, Ballone E, Lapenna D & Cuccurullo F (2009) Prognostic value of different indices of blood pressure variability in hypertensive patients. *Am J Hypertens,* 22**,** 842-7.

Poortvliet RK, Ford I, Lloyd SM, Sattar N, Mooijaart SP, de Craen AJ, Westendorp RG, Jukema JW, Packard CJ, Gussekloo J, de Ruijter W & Stott DJ (2012) Blood pressure variability and cardiovascular risk in the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER). *PLoS One,* 7**,** e52438.

Potter G, Doubal F, Jackson C, Sudlow C, Dennis M & Wardlaw J (2010) Associations of clinical stroke misclassification ('clinical-imaging dissociation') in acute ischemic stroke. *Cerebrovasc Dis,* 29**,** 395-402.

Pringle E, Phillips C, Thijs L, Davidson C, Staessen JA, de Leeuw PW, Jaaskivi M, Nachev C, Parati G, O'Brien ET, Tuomilehto J, Webster J, Bulpitt CJ, Fagard RH & Syst-Eur. investigators (2003) Systolic blood pressure variability as a risk factor for stroke and cardiovascular mortality in the elderly hypertensive population. *J Hypertens,* 21**,** 2251-7.

Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, White R, Caulfield MJ, Deanfield JE, Smeeth L, Williams B, Hingorani A & Hemingway H (2014) Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet,* 383**,** 1899-911.

Reker DM, Hamilton BB, Duncan PW, Yeh SC & Rosen A (2001) Stroke: who's counting what? *J Rehabil Res Dev,* 38**,** 281-9.

Ricci S, Celani MG & Righetti E (1994) Clinical methods for diagnostic confirmation of stroke subtypes. *Neuroepidemiology,* 13**,** 290-5.

Rinaldi R, Vignatelli L, Galeotti M, Azzimondi G & de Carolis P (2003) Accuracy of ICD-9 codes in identifying ischemic stroke in the General Hospital of Lugo di Romagna (Italy). *Neurol Sci,* 24**,** 65-9.

Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, Copper LS & Shahar E (1999) Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke,* 30**,** 736-43.

Rossignol P, Cridlig J, Lehert P, Kessler M & Zannad F (2012) Visit-to-visit blood pressure variability is a strong predictor of cardiovascular events in hemodialysis: insights from FOSIDIAL. *Hypertension,* 60**,** 339-46.

Rothwell PM (2010) Limitations of the usual blood-pressure hypothesis and importance of variability, instability, and episodic hypertension. *Lancet,* 375**,** 938-48.

Rothwell PM, Coull AJ, Giles MF, Howard SC, Silver LE, Bull LM, Gutnikov SA, Edwards P, Mant D, Sackley CM, Farmer A, Sandercock PAG, Dennis MS, Warlow CP, Bamford JM & Anslow P (2004) Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). *The Lancet,* 363**,** 1925-1933.

Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlof B, Poulter NR & Sever PS (2010a) Effects of beta blockers and calcium-channel blockers on within-individual variability in blood pressure and risk of stroke. *Lancet Neurol,* 9**,** 469-80.

Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, Sever PS & Poulter NR (2010b) Prognostic significance of visit-to-visit variability,

maximum systolic blood pressure, and episodic hypertension. *The Lancet,* 375**,** 895-905.
Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J & Griffin MR (2008) Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. *Pharmacoepidemiol Drug Saf,* 17**,** 20-6.
Ruigomez A, Martin-Merino E & Garcia Rodriguez L (2010) Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). *Pharmacoepidemiol Drug Saf,* 19**,** 579-585.
Saka O, McGuire A & Wolfe C (2009) Cost of stroke in the United Kingdom. *Age Ageing,* 38**,** 27-32.
Saver JL (2006) Optimising and Evidence-Based Causative Classification System for Ischemic Stroke. *Ann Neurol,* 59**,** 434.
Schulz UGR. & Rothwell PM (2003) Differences in Vascular Risk Factors Between Etiological Subtypes of Ischemic Stroke: Importance of Population-Based Studies. *Stroke,* 34**,** 2050-2059.
Schutte R, Thijs L, Liu YP, Asayama K, Jin Y, Odili A, Gu YM, Kuznetsova T, Jacobs L & Staessen JA (2012) Within-subject blood pressure level--not variability--predicts fatal and nonfatal outcomes in a general population. *Hypertension,* 60**,** 1138-47.
Sega R, Cesana G, Bombelli M, Grassi G, Stella ML, Zanchetti A & Mancia G (1998) Seasonal variations in home and ambulatory blood pressure in the PAMELA population. Pressione Arteriose Monitorate E Loro Associazioni. *J Hypertens,* 16**,** 1585-92.
Selvarajah JR, Glaves M, Wainwright J, Jha A, Vail A & Tyrrell PJ (2009) Classification of minor stroke: intra- and inter-observer reliability. *Cerebrovasc Dis,* 27**,** 209-14.
Sheppard JP, Holder R, Nichols L, Bray E, Hobbs FD, Mant J, Little P, Williams B, Greenfield S & McManus RJ (2014) Predicting out-of-office blood pressure level using repeated measurements in the clinic: an observational cohort study. *J Hypertens,* 32**,** 2171-8; discussion 2178.
Shimbo D, Newman JD, Aragaki AK, LaMonte MJ, Bavry AA, Allison M, Manson JE & Wassertheil-Smoller S (2012) Association between annual visit-to-visit blood pressure variability and stroke in postmenopausal women: data from the Women's Health Initiative. *Hypertension,* 60**,** 625-30.
Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM & Fried LP (2004) Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc,* 52**,** 123-7.
Sinha S, Myint P, Luben R & Khaw KT (2008) Accuracy of death certification and hospital record linkage for identification of incident stroke. *BMC Medical Research Methodology,* 8**,** 1-6.
Spolaore P, Brocco S, Fedeli U, Visentin C, Schievano E, Avossa F, Milan G, Toso V, Vanuzzo D, Pilotto L, Pessina AC & Bonita R (2005) Measuring Accuracy of Discharge Diagnoses for a Region-Wide Surveillance of Hospitalized Strokes. *Stroke,* 36**,** 1031-1034.
Springate DA, Kontopantelis E, Ashcroft DM, Olier I, Parisi R, Chamapiwa E & Reeves D (2014) ClinicalCodes: an online clinical codes repository to

improve the validity and reproducibility of research using electronic medical records. *PLoS One,* 9**,** e99825.
SSCA (2015) Scottish Stroke Care Audit (SSCA) 2015 National Report: Stroke Services in Scottish Hospitals. *ISD Scotland Publications.*
SSNAP (2015) The second Sentinel Stroke National Audit Programme (SSNAP) annual report: care received from April 2014 to March 2015. *Royal College of Physicians*.
Stegmayr B & Asplund K (1992) Measuring Stroke in the Population: Quality of Routine Statistics in Comparison with a Population-Based Stroke Registry. *Neuroepidemiology,* 11**,** 204-213.
Stuart-Buttle CD, Read JD, Sanderson HF & Sutton YM (1996) A language of health in action: Read Codes, classifications and groupings. *Proc AMIA Annu Fall Symp***,** 75-9.
Suchy-Dicey AM, Wallace ER, Mitchell SV, Aguilar M, Gottesman RF, Rice K, Kronmal R, Psaty BM & Longstreth WT (2013) Blood pressure variability and the risk of all-cause mortality, incident myocardial infarction, and incident stroke in the cardiovascular health study. *Am J Hypertens,* 26**,** 1210-7.
Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T & Collins R (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med,* 12**,** e1001779.
Sudlow CLM & Warlow CP (1996) Comparing Stroke Incidence Worldwide: What Makes Studies Comparable? *Stroke,* 27**,** 550-558.
Tai C, Sun Y, Dai N, Xu D, Chen W, Wang J, Protogerou A, van Sloten TT, Blacher J, Safar ME, Zhang Y & Xu Y(2015) Prognostic significance of visit-to-visit systolic blood pressure variability: a meta-analysis of 77,299 patients. *J Clin Hypertens (Greenwich),* 17**,** 107-15.
Teh R, Doughty R, Connolly M, Broad J, Pillai A, Wilkinson T, Edlin R, Jatrana S, Dyall L & Kerse N (2013) Agreement between self-reports and medical records of cardiovascular disease in octogenarians. *Journal of Clinical Epidemiology,* 66**,** 1135-1143.
The PLoS Medicine Editors (2012) Best Practice in Systematic Reviews: The importance of protocols and registration. *PLoS Med* 8(2): e1001009.
The Million Women Study Collaboration (1999) The Million Women Study: design and characteristics of the study population. *Breast Cancer Research,* 1**,** 1-8.
Thiru K, Hassey A & Sullivan F (2003) Systematic review of scope and quality of electronic patient record data in primary care. *BMJ,* 326**,** 1070.
Tirschwell DL & Longstreth WT (2002) Validating administrative data in stroke research. *Stroke,* 33**,** 2465-70.
Tolonen H, Salomaa V, Torppa J, Sivenius J, Immonen-Raiha P & Lehtonen A (2007) The validation of the Finnish Hospital Discharge Register and Causes of Death Register data on stroke diagnoses. *Eur J Cardiovasc Prev Rehabil,* 14**,** 380-5.
Townsend M, Wickramasinghe K, Bhatnagar P, Smolina K, Nichols M, Leal J, Luengo-Fernandez R & Rayner M. (2012) Coronary heart disease statistics 2012 edition., British Heart Foundation.

Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, Jaakkimainen L & Kapral MK (2013) Validity of Administrative Data for Identifying Patients Who Have Had a Stroke or Transient Ischemic Attack Using EMRALD as a Reference Standard. *Canadian Journal of Cardiology,* 29**,** 1388-1394.

Van Staa TP & Abenhaim L (1994) The quality of information recorded on a UK database of primary care records: A study of hospitalizations due to hypoglycemia and other conditions. *Pharmacoepidemiology and Drug Safety,* 3**,** 15-21.

Verdecchia P, Schillaci G, Reboldi G, Franklin SS & Porcellati C (2001) Different prognostic impact of 24-hour mean blood pressure and pulse pressure on stroke and coronary artery disease in essential hypertension. *Circulation,* 103**,** 2579-84.

Wahl PM, Rodgers K, Schneeweiss S, Gage BF, Butler J, Wilmer C, Nash M, Esper G, Gitlin N, Osborn N, Short LJ & Bohn RL (2010) Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf,* 19**,** 596-603.

Walker MK, Whincup PH, Shaper AG, Lennon LT & Thomson AG (1998) Validation of patient recall of doctor-diagnosed heart attack and stroke: a postal questionnaire and record review comparison. *Am J Epidemiol,* 148**,** 355-61.

Wardlaw JM, Doubal F, Armitage P, Chappell F, Carpenter T, Muñoz Maniega S, Farrall A, Sudlow C, Dennis M & Dhillon B (2009) Lacunar stroke is associated with diffuse blood-brain barrier dysfunction. *Ann Neurol,* 65**,** 194-202.

Wardlaw JM & Mielke O (2005) Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment--systematic review. *Radiology,* 235**,** 444-53.

Warlow C, Sudlow C, Dennis M, Wardlaw J & Sandercock P (2003) Stroke. *Lancet,* 362**,** 1211-24.

Warren RE, Marshall T, Padfield PL & Chrubasik S (2010) Variability of office, 24-hour ambulatory, and self-monitored blood pressure measurements. *Br J Gen Pract,* 60**,** 675-80.

Whisnant J, Basford J, Bernstein E, Cooper E, Dyken M, Easton E, Little J, Marler J, Millikan C, Petito C, Price T, Raichle M, Robertson J, Thiele B, Walker M & Zimmerman R (1990) Special report from the National Institute of Neurological Disorders and Stroke. Classification of Cerebrovascular Disorders III. *Stroke,* 21**,** 637-676.

Whiting P, Westwood M, Burke M, Sterne J & Glanville J (2008) Systematic reviews of test accuracy should search a range of databases to identify primary studies. *J Clin Epidemiol,* 61**,** 357-364.

Whiting PF, Rutjes AWF, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, MMG Leeflang, Sterne JAC & Bossuyt PMM (2011) QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine,* 155**,** 529-536.

Wolf ME, Sauer T, Alonso A & Hennerici MG (2012) Comparison of the new ASCO classification with the TOAST classification in a population with acute ischemic stroke. *J Neurol,* 259**,** 1284-9.

Wright FL, Green J, Canoy D, Cairns BJ, Balkwill A & Beral V (2012) Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. *BMC Medical Research Methodology,* 12**,** 161-161.

Wuerzner K, Hassler C & Burnier M (2003) Difficult blood pressure control: watch out for non-compliance! *Nephrology Dialysis Transplantation,* 18**,** 1969-1973.

Yamagishi K, Ikeda A, Iso H, Inoue M & Tsugane S (2009) Self-reported stroke and myocardial infarction had adequate sensitivity in a population-based prospective study JPHC (Japan Public Health Center)–based Prospective Study. *Journal of Clinical Epidemiology,* 62**,** 667-673.

Yinon L, Chen Y, Parvez F, Bangalore S, Islam T, Ahmed A, Rakibuz-Zaman M, Hasan R, Sarwar G & Ahsan H (2013) A prospective study of variability in systolic blood pressure and mortality in a rural Bangladeshi population cohort. *Prev Med,* 57**,** 807-12.

Yip PK, Jeng JS, Ng SK, Chang YC, Lee TK, Huang ZS & Chen RC (1997) The Stroke and Cerebral Atherosclerosis Study of National Taiwan University Hospital (SCAN): Background and Methodology. *Acta Neurol Taiwan,* 6**,** 300-308.

Yu JM, Kong QY, Schoenhagen P, Shen T, He YS, Wang JW, Zhao YP, Shi DN & Zhong BL (2014) The prognostic value of long-term visit-to-visit blood pressure variability on stroke in real-world practice: a dynamic cohort study in a large representative sample of Chinese hypertensive population. *Int J Cardiol,* 177**,** 995-1000.

Zhou H, Li JJ & Wang JY (2005) The reliability of ischemic stroke subtype classification using the TOAST criteria. *Chin J Intern Med,* 44**,** 825.